BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images, please do not report the images to the Image Problem Mailbox.

PCT WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification 6 :
C12N 15/54, 9/12, A61K 38/45, C12Q 1/48, C07K 16/40, G01N 33/573, A01K 67/027

A2

(11) International Publication Number: WO 96/19580

(43) International Publication Date: 27 June 1996 (27.06.96)

(21) International Application Number: PCT/US95/16531

(22) International Filing Date: 18 December 1995 (18.12.95)

(30) Priority Data:
08/359,125 19 December 1994 (19.12.94) US

(60) Parent Application or Grant
(63) Related by Continuation
US 08/359,125 (CIP)
Filed on 19 December 1994 (19.12.94)

(71) Applicant *(for all designated States except US)*: COLD SPRING HARBOR LABORATORY [US/US]; 100 Bungtown Road, Cold Spring Harbor, NY 11724 (US).

(72) Inventors; and
(75) Inventors/Applicants *(for US only)*: GREIDER, Carol [US/US]; 87 Bay Drive East, Huntington, NY 11743 (US). COLLINS, Kathleen [US/US]; 111 Prime Avenue, Huntington, NY 11743 (US). KOBAYASHI, Ryuji [JP/US]; 1152 Cove Edge Road, Syosset, NY 11791 (US). YANG, Xiaohong, Helena [CN/US]; 90 Cuba Hill Road, Greenlawn, NY 11740 (US). HEMISH, Jill, M. [US/US]; 4 Fenwood Road, Huntington Station, NY 11746 (US). AUTEXIER, Chantal [CA/US]; 188 Vernon Valley Road, East Northport, NY 11731 (US).

(74) Agents: GRANAHAN, Patricia et al.; Hamilton, Brook, Smith & Reynolds, Two Militia Drive, Lexington, MA 02173 (US).

(81) Designated States: CA, JP, MX, US, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE).

Published
Without international search report and to be republished upon receipt of that report.

U5
12
57

(54) Title: TELOMERASE PROTEIN COMPONENT

(57) Abstract

Genes encoding the amino acid sequences of the protein component of *Tetrahymena* telomerase and methods for their preparation and use are disclosed. The protein component is comprised of two subunit polypeptides, and 80 kD and a 95 kD polypeptide. Methods for deducing the protein component of human telomerase and its uses are also described. The telomerase protein component may be used to develop diagnostic procedures for detection of telomerase activity in cancers, microbial diseases, and other disorders, and to produce therapeutic compounds to inhibit or enhance telomerase activity.

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	**GB**	United Kingdom	**MR**	Mauritania
AU	Australia	**GE**	Georgia	**MW**	Malawi
BB	Barbados	**GN**	Guinea	**NE**	Niger
BE	Belgium	**GR**	Greece	**NL**	Netherlands
BF	Burkina Faso	**HU**	Hungary	**NO**	Norway
BG	Bulgaria	**IE**	Ireland	**NZ**	New Zealand
BJ	Benin	**IT**	Italy	**PL**	Poland
BR	Brazil	**JP**	Japan	**PT**	Portugal
BY	Belarus	**KE**	Kenya	**RO**	Romania
CA	Canada	**KG**	Kyrgystan	**RU**	Russian Federation
CF	Central African Republic	**KP**	Democratic People's Republic of Korea	**SD**	Sudan
CG	Congo			**SE**	Sweden
CH	Switzerland	**KR**	Republic of Korea	**SI**	Slovenia
CI	Côte d'Ivoire	**KZ**	Kazakhstan	**SK**	Slovakia
CM	Cameroon	**LI**	Liechtenstein	**SN**	Senegal
CN	China	**LK**	Sri Lanka	**TD**	Chad
CS	Czechoslovakia	**LU**	Luxembourg	**TG**	Togo
CZ	Czech Republic	**LV**	Latvia	**TJ**	Tajikistan
DE	Germany	**MC**	Monaco	**TT**	Trinidad and Tobago
DK	Denmark	**MD**	Republic of Moldova	**UA**	Ukraine
ES	Spain	**MG**	Madagascar	**US**	United States of America
FI	Finland	**ML**	Mali	**UZ**	Uzbekistan
FR	France	**MN**	Mongolia	**VN**	Viet Nam
GA	Gabon				

Telomerase Protein Component

Background of the Invention

Chromosome stability is essential for cell viability. Eukaryotes have linear chromosomes and the telomeres that cap the ends protect chromosomes from degradation and recombination. Loss of telomeric DNA during cell proliferation may play a role in ageing and cancer. Counter, C.M., *et al.* (1992) *EMBO, 11*:1921-1929.

Telomeric sequences are highly conserved in eukaryotes. The DNA sequence contains simple tandem repeats of specific GT-rich motifs. The exact sequences are characteristic of a particular organism; i.e., d(TTGGGG) in *Tetrahymena*, d(TTTTGGGG) in *Oxytricha* and d(TTAGGG) in humans. The number of repeats on any given chromosome end varies, giving telomeres a characteristic heterogeneous or "fuzzy" appearance on Southern blots. In addition to sequence conservation, telomere function is also conserved in eukaryotes. *Tetrahymena* and human telomeres function in the yeast *Saccharomyces cerevisiae*, and yeast telomeres function in other fungi. Szostak, J.W. and E.H. Blackburn (1982) *Cell, 29*:245-255. Thus the mechanisms for maintaining a stable end must share essential features in diverse eukaryotes.

Telomere sequences are synthesized onto chromosome ends by a highly specialized DNA polymerase called telomerase. Telomerase is a ribonucleoprotein enzyme in which both the RNA and the protein components are essential for telomerase activity. The RNA component provides the template for the telomere repeat synthesis. Blackburn, E.H. (1992) *Annu. Rev. Biochem., 61*:113-129.

Telomere replication involves the establishment of an equilibrium between telomere shortening and telomere

lengthening. DNA replication leads to telomere shortening because DNA-t mplate dependent DNA polymerase cannot replicate the very end of a DNA molecule. Telomerase elongates chromosomes through *de novo* sequence addition. Double-stranded synthesis by DNA-template dependent DNA polymerase and primers then fill in the complementary C-rich strand.

The RNA component of telomerase has been sequenced for humans (Feng, J., *et al.* (1995) *Science 269*:1236-1241), mice, and several mammalian species (Greider, C., unpublished data), as well as *Saccharomyces cerevisiae, Tetrahymena, Euplotes* and *Oxytricha*. See Singer and Gottschling, (1994) *Science, 266*:404-409; Lingner, *et al. (1994) Genes & Development, 8*:1984-1988; Romero, D.P. and E.H. Blackburn (1994) *Cell, 67*:343-353. The protein component of a telomerase from any species has not previously been sequenced or cloned.

Summary of the Invention

Described herein are genes encoding a telomerase protein component of eukaryotic, including mammalian, origin, telomerase proteins encoded by the genes, RNA encoding the polypeptides described, and sequences that hybridize to these genes. As described herein, the genomic sequences encoding a telomerase protein component have been determined by the Applicants. Both the RNA and the protein components of telomerase are essential in the maintenance of telomeric length in chromosomes. The protein component of a telomerase can be used by itself or coupled with the RNA component in diagnostic or therapeutic methods and in assays for telomerase.

As described herein, a *Tetrahymena* gene encoding a *Tetrahymena* telomerase protein component has been isolated and sequenced. The polypeptides encoded by these genes have been shown to be an 80 kD and a 95 kD polypeptide (p80

and p95, respectively). The polypeptides comprise a protein that, coupled with the RNA component, acts to add telomeric TTGGGG repeats to stabilize chromosomal telomere length. The present invention also provides DNA sequences and portions thereof, sequences complementary to these DNA sequences, and sequences, such as probes, that hybridize to either the sense or the complementary (antisense) sequences or fragments thereof that encode the polypeptides disclosed.

In particular, an 80 kD and a 95 kD polypeptide which are components of *Tetrahymena* telomerase protein have been isolated and sequenced. The amino acid sequences of the 80 kD and 95 kD polypeptides of the protein component are disclosed herein, as are the DNA (nucleic acid) sequences which encode the 80 kD and 95 kD proteins.

Further disclosed are nucleotide sequences encoding p80 and p95 telomerase polypeptides which are translated by most eukaryotes. These DNA sequences have been incorporated into plasmids, and the plasmids transfected into vectors. Host cells comprising these vectors are provided for the production of recombinant telomerase protein component.

Both DNA sequences and polypeptide sequences that are substantially equivalent to the disclosed sequences are also provided by this invention.

Also included are methods of using the DNA sequences encoding the *Tetrahymena* protein components to determine the DNA sequences encoding the protein components of other invertebrate and vertebrate species, in particular mammalian species, such as the genes for human, mouse, rat, dog, cat, pig, chimpanzee, or monkey telomerase protein component.

The present work also makes available methods of determining whether a mammal, especially a human individual, is likely to be affected with a disorder or

disease in which abnormal telomerase activity is a symptom or cause. Methods of detecting telomerase expression are provided as a means of diagnosing a predisposition to the development of immortal or cancer cells in a human or in another animal. In one embodiment, DNA or RNA present in a cell or tissue sample is hybridized to a DNA or RNA probe which is complementary to all or a portion of a telomerase protein component gene. As used herein, the term "telomerase protein component gene" includes the genes whose sequence is described herein, genes which hybridize to the genes or portions thereof, and equivalent genes from other species, such as those from human, mouse, rat, dog, cat, pig, chimpanzee, monkey, or *Tetrahymena*. Detection of hybridization is an indication of a predisposition to the development of or the presence of cancer, or another disorder in which immortal cells arise.

An important feature of this invention is that the telomerase protein component can be used to screen for telomerase inhibitors which can be used to prevent telomerase expression and/or activity in cells. The protein component can be used as a basis for a method to identify and treat individuals affected by abnormal telomerase activity either within their own cells and tissues, or in foreign cells of invading parasites or disease organisms which are eukaryotes.

Therefore, the present invention provides a diagnostic tool through which inhibitors of telomerase activity can be tested and developed, and by which diseases such as cancer, or infections, such as yeast or protozoan diseases, can be diagnosed.

Another embodiment of the present invention is antibodies to a telomerase protein component, such as antibodies to one or both of the 80 kD or 95kD polypeptides, synthetic telomerase polypeptide sequences, or portions of these polypeptides. These include both

polyclonal and monoclonal antibodies, such as polyclonal and monoclonal antibodies which bind either or both the 80 kD and 95 kD polypeptides of this invention. Such anti-telomerase antibodies are useful to detect telomerase activity in cells and tissues.

Further embodiments include methods of therapy and treatment involving recombinant and/or transgenic cells containing either or both of the genes for the telomerase subunits, by themselves or in combination with other genes, such as a gene encoding the telomerase RNA component. Recombinant or transgenic cells producing anti-telomerase antibodies are included as well. Such methods can be applied to the treatment of disorders arising from abnormal telomerase activity or can be used to increase or trigger expression of telomerase to prevent cell mortality.

Brief Description of the Figures

Figure 1 is the nucleotide sequence (SEQ ID NO:1) of the *Tetrahymena* 80 kD protein gene. The nucleotide sequence is derived from genomic and cDNA clones.

Figure 2 is the amino acid sequence (SEQ ID NO:2) of the 80 kD protein deduced from the nucleotide sequence shown in Figure 1.

Figure 3 is the nucleotide sequence (SEQ ID NO:3) of the *Tetrahymena* 95 kD protein gene.

Figure 4 is the amino acid sequence (SEQ ID NO:4) of the 95 kD protein deduced from the nucleotide sequence shown in Figure 3.

Figure 5 is primer set 1 consisting of 12 deoxyribonucleotide sequences (primers 1-12).

Figure 6 is primer set 2 consisting of 12 deoxyribonucleotide sequences (primers 13-24).

Figure 7 is primer set 3 consisting of 10 deoxyribonucleotide sequences (primers 25-34).

Figures 8A-8B show primer set 4 consisting of 18 deoxyribonucleotide sequences (primers F1-F9 and R1-R9).

Figure 9 is primer set 5 consisting of 10 deoxyribonucleotide sequences (primers F10-F14 and R10-R14).

Figure 10 is the DNA sequence (SEQ ID NO:8) of the genetically-engineered p80 gene.

Figure 11 is the DNA sequence (SEQ ID NO:9) of the genetically-engineered p95 gene.

Detailed Description of the Invention

This invention relates to genes encoding a eukaryotic telomerase protein component, the polypeptides encoded by these genes, as well as the RNA encoding the polypeptides, complementary nucleotide sequences, and probes that hybridize to sense and complementary portions of the nucleotide sequences.

Further provided are synthesized genes encoding 80 kD and 95 kD telomerase protein components, the recombinant polypeptides encoded by these genes, the RNA encoding the polypeptides, the primers used to synthesize these genes, and complementary nucleotide sequences or fragments thereof.

Those skilled in the art will appreciate that many different DNA sequences can encode a single protein. In addition to the genes and other nucleotide sequences described above, contemplated within this invention are DNA sequences which encode catalytically active, telomerase protein components, and nucleotide sequences that hybridize to these DNA sequences. Generally, these will hybridize under moderately stringent conditions. According to the invention, the term "stringent conditions" means hybridization conditions comprising a salt concentration of 4X SSC (NaCl-citrate buffer) at 62° -66° C., and "high

stringent conditions" means hybridization conditions comprising a salt concentration of 0.1X SSC at 68° C. Ausubel, *et al.*, (1994) *Current Protocols in Molecular Biology*, John Wiley & Sons, Inc.

Methods of using these sequences to deduce other telomerase components are described. Methods of diagnosis and treatment which use a telomerase protein component, nucleotide sequences encoding the protein component or portions thereof are also included.

Following is a description of the embodiments of the invention, which, together with the following examples delineating the experimental procedures, serve to explain the principles of the invention. All references to materials and methods are herein incorporated by reference.

The present invention also encompasses polypeptides comprising a telomerase protein component of eukaryotic origin, including the polypeptides herein described. All polypeptides which comprise a telomerase protein component and are active as a component of a telomerase are encompassed by the present invention and the term telomerase protein component as used herein.

The telomerase protein component has been produced by the following method in "substantially pure" form. "Substantially pure" is defined as the minimum amino acid sequence that, when combined with the telomerase RNA component, demonstrates telomerase activity.

Tetrahymena Protein Component Genes

Tetrahymena telomerase enzyme was purified using readily available chromatography matrixes. Two criteria were used to follow enzyme purification. First, activity assays were performed using the standard telomerase assay (Greider, C.W. (1987) *Cell*, *51*: 887-898) and 32P-dGTP incorporation was quantitated by spotting on DE-81 paper

and determining the counts incorporated (Greider, C.W. (1987) Ph.D. Thesis, Univ. Calif. Berkeley). Second, telomerase RNA was followed by Northern blot analysis and quantitated by comparison to a titration of a known amount of a synthetic telomerase RNA standard. Purification over hydroxylapatite, spermine agarose, Sepharose CL-6B sizing column, phenyl-Sepharose, DEAE agarose (or Q-Sepharose) and a 15- or 20-35% glycerol gradient, yielded highly purified telomerase fractions. Two predominant proteins of 80 and 95 kD were identified in the active fractions which co-purified with telomerase activity and were present in a stoichiometry similar to the telomerase RNA.

Two samples of the material purified as described above were separated on a non-denaturing gel. One lane of the gel was Northern blotted to identify the position of the telomerase RNA and the other lane was cut from the native gel and run in a second dimension on an SDS PAGE gel. Most of the proteins remained near the well of the first native gel; however, both the telomerase RNA and the p80 and p95 proteins ran approximately one-third of the way into the native gel at equivalent positions, indicating that p80 and p95 are components of telomerase. Beginning with over 300 L or $1.2x10^{11}$ cells, the active fraction in the final glycerol gradient contained over a microgram of telomerase RNA. This indicated there was enough material to sequence the co-purifying polypeptides.

To determine if the p80 and p95 fraction comprises telomerase activity or is a contaminant that migrates with the same properties, telomerase was treated with micrococcal nuclease. Previous experiments have shown that limited cleavage of the telomerase RNA does not completely inactivate telomerase activity. Greider, C.W. and E.H. Blackburn (1989) *Nature*, 337:331-337. Two fractions of purified telomerase were prepared for glycerol gradient

analysis, to determine whether cleavage of the RNA would alter the mobility of the RNP in a glycerol gradient. One sample was briefly treated with micrococcal nuclease; the other was incubated with buffer only. These samples were sedimented through a glycerol gradient and fractions were collected from each gradient. The activity was assayed and the protein profile determined. In the untreated fraction, activity peaked in fractions 8 and 9 along with p80 and p95. In the micrococcal nuclease-treated fraction weak activity peaked in fraction 10, the peak of p80 and p95 was now also shifted to fraction 10, indicating that these proteins behave as expected for telomerase components. The sedimentation of most other proteins in the gradient remained unchanged relative to the change in sedimentation of telomerase.

The partial peptide sequences from both p80 and p95 were determined. The complete amino acid sequences of the two polypeptides can be determined in the same manner. Telomerase from 344 L of Tetrahymena cells was purified according to the procedures described above with the addition of a DEAE agarose concentration step followed by non-denaturing gel electrophoresis and SDS PAGE electrophoresis. To avoid problems associated with direct N-terminal sequencing of proteins, the excised protein bands were digested with Lysylendopeptidase from *Achromobacter*. The peptide fragments were extracted from the gel and resolved on a C18 reverse phase HPLC column. Several well defined peptide peaks were subjected to successive rounds of Edman degradation on an Applied Biosystems automated sequencer. From two separate preparations of telomerase, the amino acid sequence was determined for 7 peptides from p80 and for 25 peptides from p95. Degenerate oligonucleotides were designed using the *Tetrahymena* codon bias as a guide. Martindale, D.W. (1989) *J. Protozol. 36:29-34*.

Oligonucleotides w re used in sets of two to obtain PCR products from either reverse transcribed RNA or from genomic DNA. Two PCR products were obtained for each protein gene. The sequence of three of the four PCR products encoded peptides which had been identified by protein sequencing but were not used as primers for the PCR (Figures 2 and 4). Genomic Southern blots probed with the PCR products for either p80 or p95 proteins showed that the gene probably exists as a single copy in the Tetrahymena genome. Northern blot analysis from actively growing cells showed a single band of about 3.0 kb for p95 and a single band of 2.5 kb for the p80 mRNA. RNA from stationary cells showed two bands when probed with the 5′ portion of the 95 kD gene. This suggests alternative processing of this gene.

To obtain the full length protein sequence, the cloned PCR products were used as probes for both *Tetrahymena* cDNA libraries and genomic libraries. Positive clones were obtained, subcloned, and sequenced. To deduce these protein sequences, the *Tetrahymena* genetic code was used since this sequence differs from that of other eukaryotes. Prescott, D.M. (1994) *Microbiol. Rev., 58*:233-267. Applicants have determined the sequence for the entire open reading frame (ORF) for both the p80 and p95 proteins (Figure 1, SEQ ID NO:1 and Figure 3, SEQ ID NO:3, respectively). The nucleotide sequence is derived from genomic and cDNA clones; polyadenylation of the mRNA occurs near the 3′ end of the reported sequence. First, Northern blot analysis of the p80 and p95 mRNAs suggests sizes of approximately 2.47 and 2.9 kb. Applicants have obtained more than 2.4 and 2.8 kb of sequence for these mRNAs. Second, all reliable peptide sequence was found in the ORFs (7/7 for p80; 25/25 for p95). Third, the suggested translation of the mRNAs from the first methionine codon in the longest ORF yields predicted protein products of equal

or slightly greater molecular mass than predicted from analysis of the proteins by SDS-PAGE. Fourth, sequences outside the region translated as coding contain a higher content of A/T than coding regions, typical of Tetrahymena genes. Prescott, D.M. (1994) *Microbiol. Rev.*, *58*:233-267. Neither of the genes has a counterpart in Genbank, EKMBL, PIR and Swissprot databases.

To demonstrate that the 80 kD and 95 kD proteins are components of telomerase, polyclonal antibodies were generated against the two proteins. Synthetic peptides were synthesized that corresponded to two different regions from each protein. Two polyclonal antibodies to peptides of the 80 kD protein (designated A81 and A82) and four antibodies to peptides of the 95 kD protein (designated A83, A84, A85, and A86) showed good titre against the respective proteins.

Table 1 lists the antibodies obtained to various peptide sequences used. The peptide sequence list was obtained directly from protein sequencing of PCR products. The first peptide was derived from a preliminary sequencing trial and was determined to be incorrect after the gene was cloned. This peptide and the antibodies directed against it were subsequently used as controls. The peptide injected into rabbits to produce A85 and A86 has one error (a missing T at the penultimate position) relative to the cloned sequence; however, the antibodies against this peptide cross react with the 95 kD protein. An N-terminal C residue was added to each peptide during synthesis in order to couple the peptide to carrier protein.

TABLE 1

Antibodies Generated Against Peptide Sequences

Antibody #	Protein Directed Against	Peptide Sequence	Amino Acid # in Sequence
79		poor (incorrect) sequence	
80		poor (incorrect) sequence	
81	80kD	(C)AEGYSDINVRG	628-
82	80kD	(C)AEGYSDINVRG	628-
83	95kD	(C)QNEFQFNNVK	610-
84	95kD	(C)QNEFQFNNVK	610-
85	95kD	(C)EFGLEPNILK	414-
86	95kD	(C)EFGLEPNILK	414-

Of these antibodies, those with the highest affinity for the 80 kD protein (A82) and the 95 kD protein (A86) were used to demonstrate that both the 80 kD and 95 kD polypeptides co-purified with telomerase activity, indicating that these proteins are telomerase components. Results of immunoprecipitation studies with the 80 kD protein are consistent and suggest the 80 kD protein is a functional component of telomerase.

Synthetic Protein Component Genes

To produce genes that encode the telomerase protein components in other eukaryotes, synthetic gene sequences were constructed in which the *Tetrahymena* genetic code was altered to enable correct translation and transcription in organisms having the genetic code which is used/translated in most eukaryotes; *i.e.*, mammals such as humans. The

genes were synthesized as gene fragments from overlapping sets of oligonucleotides (primer sets), which were then cloned into plasmids. The full-length genes were constructed by combining the fragments in the plasmids. The p80 gene was constructed in the plasmid Bluescript; the p95 gene in the plasmid pSE280, although any plasmid can be used.

To express the p80 and p95 proteins, the synthesized genes were cloned into different restriction sites of the pRSET and pBlueBac vectors. Transcription and translation of the genes in PRSET and pBlueBac generates the recombinant proteins in *E. coli* and baculovirus, respectively. A His-tag and cleavage site at the end of each recombinant protein facilitates the purification of the proteins.

Using *E. coli* and a vector such as pRSET containing the p80 and p95 gene constructs or, alternatively, baculovirus and a vector such as pBlueBac containing the same constructs, p80 and p95 can be expressed recombinantly. Thus, applicants have produced the first known bacterial strains or expression vectors which permit expression of the p80 and p95 telomerase protein components. One embodiment of this invention is the production of one or more recombinant telomerase protein components in a host cell. One method comprises culturing a host cell containing the gene encoding the protein, or a homolog thereof, under conditions which permit production of the protein. In one embodiment, the method further comprises the steps of recovering quantities of protein, as well as purification procedures. Skilled artisans will appreciate the various ways in which recombinant proteins of this invention can be prepared.

The recombinant proteins, or fragments thereof, are useful to detect agents that stimulate or inhibit telomerase catalytic activity. They are also useful to

produce antibodies for screening assays, such as to detect telomerase activity in tumor cells, or stimulated or inhibited production of telomerase in response to exposure to a compound.

Also encompassed by this invention is a telomerase polypeptide which varies in amino acid sequence from a telomerase polypeptide encoded by genomic DNA (*i.e.*, differs from a naturally-occurring telomerase polypeptide, such as *Tetrahymena* telomerase polypeptide), without affecting the ability of the polypeptide to combine with the other telomerase protein and RNA components or affect the enzymatic activity of telomerase. These variations may include additions, deletions, substitutions and other alterations (*e.g.*, modification of an amino acid residue) to the amino acid sequences.

The genes encoding the *Tetrahymena* telomerase protein component, the synthesized genes, or the primers can be used to clone the human telomerase protein component and other mammalian telomerase protein components, using known methods described herein.

Two approaches can be used to clone the human telomerase protein genes with the *Tetrahymena*, synthesized, or primer sequences. These procedures are described in detail in the subsequent examples.

In one approach, DNA sequence hybridization is used to identify and clone a human homologue of the *Tetrahymena* protein. Human genomic DNA and mRNA blots are probed with the *Tetrahymena* gene at a series of increasing stringencies. If specific bands are identified, cDNA or genomic libraries cloned into phage lambda vectors are probed at a similar stringency to identify the gene for the human homologue. Positive phage are restriction mapped, subcloned and sequenced.

A second approach is to produce a series of antibodies to various regions of both the p80 and p95 proteins. The

antibodies described above (A62 and A86) can be used. Additional antibodies can also be generated as synthetic peptides and fusion proteins and used to identify human telomerase proteins by cross-reactivity. Libraries of human or other mammalian cDNAs which express a portion of the protein can be probed with the antibodies to clone the human or mammalian genes by standard molecular biology procedures. See, Sambrook, *et al.* (1989) *Molecular Cloning - A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor Laboratory, NY.

Human telomerase is an excellent target for anti-cancer therapy. The availability of the protein components for the *Tetrahymena* enzyme facilitates a thorough understanding of telomerase biochemistry and will aid in the identification of specific anti-telomerase drugs.

Telomerase activity has been found in over 70 immortalized human cell lines and cancer tissues, but few human primary somatic cells or tissues. Kim, *et al.* (1994) *Science 266*:2011-2015. Telomere length maintenance does not occur in primary human somatic cells that have a limited life span. When primary cells divide, either *in vitro* or *in vivo*, telomere length shortens. Germline cells do not show this shortening. Allsopp, *et al.* (1992) *PNAS 89*:10114-10118; Harley, *et al.* (1990) *Nature, 337*:331-337; Vaziri, *et al.* (1993) *Amer. J. Hum. Genet. 52*:661-667. Although telomerase activity is present in immortalized human HeLa cells (Morin, G.B. (1989) *Cell 59*:521-529), telomerase has not been detected in primary fibroblast cultures. Applicants established SV40 immortalized lines from primary human cells to investigate the connection between telomerase and telomere shortening. In both primary and SV40 transfected human embryonic kidney cells, telomeres shortened but telomerase was not detected as the cells were passaged. When the culture underwent crisis most cell lines died; however, in the immortal clones that

survived, telomerase activity was detected and the telomeres were short but stably maintained. Counter, *et al.* (1992) *EMBO J.* *11*:1921-1929. Similar results were obtained with primary mouse cells in culture. Prowse, K.R. and C.W. Greider (1995) *PNAS* *92*:4818.

These results suggest that primary cells express little or no telomerase activity, but that following immortalization, cancer cells reactivate telomerase and maintain telomere length. In fact, telomerase activity has been demonstrated in human ovarian carcinoma cells, but not in normal cervical endothelial cells. Counter, *et al.* (1994) *PNAS* *91*:2900-2904. Telomere shortening before crisis may be lethal, but those cells that can reactivate telomerase maintain telomere length and survive crisis. This model suggests that if telomerase is required for the growth of immortalized cells, telomerase inhibitors may be excellent anti-cancer drugs.

The present work provides a method by which cancers may be diagnosed prior to or during clinical manifestation of symptoms by means of detecting telomerase activity in somatic cells that normally do not express telomerase. Telomerase mRNA expression in a sample of somatic cells or tissue can be detected using DNA or RNA probes; this is indicative of expression of telomerase which, in turn, is an indication of immortal cancer cells since somatic cells do not normally produce telomerase. Detection of hybridization is an indication of a predisposition to cellular immortalization or cancer, or to the presence of cancer or immortal cells.

By hybridization, it is meant that DNA and/or RNA molecules or portions thereof are used in a hybridization analysis to detect complementary polynucleotides under conditions of moderate stringency according to methods described in Ausubel, *et al.*, (1994) *Current Protocols in Molecular Biology*, (Suppl. 26), John Wiley & Sons, Inc.

In one embodiment of detecting the presence of immortal cells or a predisposition to immortalization in a eukaryotic tissue sample or a sample of eukaryotic cells, nucleic acids are used as probes or primers. This embodiment may comprise the steps of:

a) obtaining a tissue sample or a sample of cells from the eukaryote; and

b) determining the presence of telomerase in the sample, wherein if the sample demonstrates the presence of telomerase, immortal cells or the predisposition to immortalization is present. The same method may be used to detect a predisposition to cancer or the presence of cancer cells or tissue.

Alternatively, the expression of mammalian telomerase can be detected using polyclonal or monoclonal antibodies to the p80 or p95 polypeptide subunits, to both subunits or fragments thereof. An antibody can detect both subunits or two antibodies can be used, each of which detects a different subunit. For example, a sample of somatic or tumor cells from an individual can be contacted with anti-telomerase antibodies after the sample has been processed or treated to render the telomerase (if present) available for binding to the antibody. Binding of the antibody is indicative of the presence of telomerase and, thus, an indication of cellular immortalization or a predisposition to cancer, or the presence of cancer or immortal cells.

A method using antibodies to detect telomerase in a eukaryotic tissue sample or a sample of eukaryotic cells may comprise the steps of:

a) obtaining a tissue sample or a sample of cells from the eukaryote; and

b) treating the sample to render telomerase available for binding to anti-telomerase antibodies, thereby producing a treated sample;

c) contacting the treated sample with anti-telomerase antibodies against (polyclonal or monoclonal) telomerase; and

d) detecting binding of the antibodies to telomerase, wherein if binding occurs, telomerase is present. It will be appreciated that antibody detection can be useful not only to detect cellular immortalization such as occurs with the development of cancer cells, and the presence of cancer or immortal cells, but also to detect the presence of foreign eukaryotic cells in the cells and tissues of a multicellular organism, as described below.

The present invention also provides a means for developing drugs and pharmaceutical compounds that destroy or otherwise inactivate or interfere with the activity of telomerase. A compound that inhibits or inactivates *Tetrahymena* telomerase activity can also be assessed for its effects on mammalian telomerases. The telomerase protein component, either with or without the RNA component, can be used to screen for drugs and pharmaceutical compounds effective as anti-cancer and anti-microbial agents, as described below.

Further, since additional telomerase activity may have an anti-aging effect and result in restoration of cells by stabilizing telomere length, compounds can be screened for their ability to stimulate or trigger telomerase activity. The protein components can also be combined with the RNA component of telomerase to produce a functional telomerase molecule which can be delivered to cells by conventional methods. Alternatively, DNA encoding a telomerase molecule can be introduced into target cells by recombinant DNA methods and transformation technology. The incorporation of extra copies of functional telomerase molecules may extend the replicative life span of the host cell by stabilizing telomere length. Thus, this invention includes methods for gene therapy in mammals.

Another application of this invention is the detection of eukaryotic disease-causing organisms in somatic cells and tissues of mammals and treatment of the resulting disease. There are many fungi, protozoa, and even algae that invade the cells and tissues of mammals and are the cause of various diseases. Examples of such diseases include, but are not limited to, aspergillosis, histoplasmosis, candidiasis, paracoccidioidomycosis, malaria, trichinosis, filariasis, trypanosomiasis (sleeping sickness), schistosomiasis, toxoplasmosis, and leishmaniasis. These organisms require telomerase and express this enzyme as they multiply inside host cells which do not normally produce telomerase. The above-described methods to detect telomerase can be used to develop early detection and diagnosis procedures for these eukaryotic microbial parasites.

An example of such a method to detect a disease caused by a eukaryotic microbial organism in a tissue sample or a sample of eukaryotic cells from an individual may comprise the steps of:

a) obtaining a tissue sample or a sample of cells from the individual; and

b) determining the telomerase in the sample, wherein if the sample demonstrates telomerase of a eukaryotic microbe, a disease caused by a eukaryotic microbial organism is present.

The telomerase in the sample can be determined by the use of nucleic acid probes or primers, including, but not limited to those described herein; or, by the use of antibodies which bind to a telomerase protein component.

Furthermore, since mammalian somatic cells do not require telomerase, the use of inhibitors of and antibiotics against telomerase will provide a method of treatment for such diseases that is nontoxic or exhibits little toxicity to the host. For example, most of the

drugs used to treat diseases caused by *Trypanosoma* species can cause serious side effects and even death. Antisense RNA to the 80 kD or 95 kD protein component of *Trypanosoma* sp. telomerase or drugs against telomerase can be used to inhibit telomerase and thus prevent the multiplication of species of this parasite in an individual without affecting the host's somatic cells and tissues. Included among these pharmaceuticals are antisense nucleic acids that inhibit the translation of mRNA encoding the protein component of telomerase.

Compounds that inhibit or destroy telomerase activity can be formulated into pharmaceutical compositions containing a pharmaceutically acceptable carrier and/or other excipients using conventional materials and means. They can be administered to an animal, either human or non-human, for therapy of a disease or condition resulting from an abnormal level of telomerase activity. Administration may be by any conventional route (parenteral, oral, inhalation, and the like) using appropriate formulations, many of which are well known. The compounds can be employed in admixture with conventional excipients, such as pharmaceutically acceptable organic or inorganic carrier substances suitable for parenteral administration that do not deleteriously react with the active derivatives.

It will be appreciated that the actual preferred amounts of active compound in a specific case will vary according to the specific compound being utilized, the particular compositions formulated, the mode of application, the particular situs of application, and the individual being treated. Dosages for a given recipient will be determined on the basis of individual characteristics, such as body size, weight, age and the type and severity of the condition being treated.

It should be noted that the formulations described herein may be used for veterinary as well as human

applications and that the term "individual" or "host" should not be construed in a limiting manner. These terms include human and nonhuman vertebrates, particularly mammals.

In a further aspect, the present invention provides a process for producing a recombinant product comprising:
(a) producing an expression vector which includes DNA which encodes a telomerase molecule;
(b) transfecting or infecting a host cell with the vector; and
(c) culturing the transfected or infected cell line to produce the encoded telomerase molecule (recombinant telomerase). The standard techniques of molecular biology can be used to prepare DNA sequences coding for the RNA and protein components of telomerase, and for construction of vectors with appropriate promoters for enzyme expression in a host cell. Suitable host cell/vector systems, transfection or infection methods and culture methods are well known in the art. These systems may also be used to produce antibodies to telomerase.

It will also be appreciated that the methods described above may be used to produce transgenic cells, tissues, and organisms for use in investigating the role of telomerase in eukaryotic organisms, and for therapeutic purposes. Thus, this invention provides transgenic biological materials that comprise the protein components of telomerase from eukaryotes, including mammals.

The present invention will now be illustrated by the following examples, which are not intended to be limiting in any way.

Example 1

Purification of Telomerase

Tetrahymena thermophila were grown to a density of 4.0 x 10^5/ml in PPYS (2% proteose peptone, 0.2% yeast extract, 10 μM $FeCl_3$), harvested by centrifugation in a GSA rotor (Sorvall), and starved for 18 h in Dryls (1.7 mM $NaC_6H_8O_7$, 1.2 mM NaH_2PO_4, 1.3 mM Na_2HPO_4, 2 mM $CaCl_2$). Starved cells were again harvested, resuspended in T2MG buffer (20 mM Tris-HCl pH 8.0, 1 mM $MgCl_2$, 10% glycerol, 2 mM DTT or β-mercaptoethanol, 0.1 mM PMSF, 2 μg/ml leupeptin, 1 μg/ml pepstatin), and lysed by addition of a final concentration of 0.2% NP-40. S-100 extract was obtained by centrifugation of lysed cells at 130,000 x g for 50 min at 4oC. All subsequent steps were done at 4oC.

S-100 extract derived from 1-2x10^{11} cells (approximately 300 L of PPYS culture) was filtered coarsely and applied to ceramic HAP (AIC) equilibrated in T2MG. Telomerase was eluted with a gradient from 0.2 M K_2HPO_4 in T2MG. Fractions with peak activity were pooled, diluted with 3 volumes of T2MG, and applied to Spermine agarose (Sigma) equilibrated in T2MG with 0.15 M potassium glutamate ($KC_5H_8NO_4$, abbreviated KG). Telomerase was eluted in T2MG with 0.65 M KG. Fractions with peak activity were pooled and loaded on a 1 L column of Sepharose CL-6B (Pharmacia) equilibrated and run in T2MG with 20 mM KG and 3 mM NaN_3. Fractions with peak activity were pooled, adjusted to 0.4 M KG, and applied to Phenyl Sepharose (Pharmacia) equilibrated in T2MG with 0.4 M KG. The column was washed in T2MG, then telomerase was eluted in T2MG with 1% Triton X-100. Fractions with peak activity were pooled and applied to DEAE agarose (BioRad) equilibrated in T2MG. Telomerase was eluted with a gradient or a step to 0.4 M KG in T2MG. Fractions with peak activity were sometimes

diluted with distilled water and were layered on 15- or 20-35% glycerol gradients. Gradients were centrifuged for 20 h in an SW41 rotor (Beckman). Glycerol gradient-purified telomerase was used in several experiments described in this invention.

Telomerase was additionally purified prior to proteolytic digests for peptide sequencing. Glycerol gradient fractions of peak activity were pooled and applied to DEAE agarose equilibrated in T2MG. Telomerase was eluted in T2MG with 0.4 M KG. Peak fractions were dialyzed against T2MG then applied to a 6% acrylamide, 50 mM Tris-acetate gel run in 50 mM Tris-acetate buffer, pH 8.0. The native gel was run for approximately 12 h at approximately 250 V. The native gel lane containing telomerase was excised, soaked briefly in 2X SDS sample buffer (0.125 M Tris-HCl pH6.8, 4% SDS, 10% β-mercaptoethanol, 20% glycerol, bromophenol blue) and sealed into the well of a denaturing 7% acrylamide gel with 0.1% agarose in mM Tris-acetate. SDS-PAGE was performed in Tris-glycine-SDS buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS).

Example 2

Analytical Scale Two-Dimensional Gel Analysis

Fractions were adjusted to at least 10% glycerol and loaded on a native gel of 6% acrylamide, 50 mM Tris-acetate minigel. The native gel was run in 50 mM Tris-acetate buffer, pH 8.0. The native gel lane containing telomerase was excised, soaked briefly in 2X SDS sample buffer, and sealed into the well of a denaturing 5-15% or 5-20% gradient acrylamide minigel with 0.1% agarose in 25 mM Tris-acetate. SDS-PAGE was performed in Tris-glycine-SDS buffer (25 mM Tris-HCl pH 8.3, 192 mM glycine, 0.1% SDS). After electrophoresis, gels were soaked 2 x 10 min in 50% methanol and equilibrated in 5% methanol. Silver staining was performed by incubation of the gel in 0.1 mM DTT for 20

min, 1 mg/ml silver nitrate for 20 min, and development in 0.28 M Na_2CO_3, 0.0185% formaldehyde. The staining reaction was quenched with citric acid.

Example 3

Proteolytic Digestion of Telomerase Subunits and Peptide Sequencing

The 80 and 95 kD subunits of telomerase were purified as described above. Preparative SDS gels containing telomerase were stained in 0.05% Coomassie brilliant blue (Aldrich), 20% methanol, 0.5% acetic acid and destained in methanol-acetic acid. Polypeptides were excised from the gel after soaking 10 min in distilled water. Gel slices were crushed and soaked in 50% methanol 2 x 20 min, decanted, and dried briefly under vacuum. Proteins were digested with approximately 300 ng of *Achromobacter* protease I in 0.1 M Tris-HCl pH 9.0, 0.01% Tween-20 for 24 h at 37°C. Peptides were separated from gel fragments by spin filtration, concentrated by Speed-Vac, and applied to a C-18 column (Vydac). Peptides were eluted with a gradient of acetonitrile:isopropanol (3:1) in 0.09% trifluoroacetic acid. Peaks of absorbance at 214 nm were collected, lyophilized, and applied to a protein sequencer (ABI).

Example 4

Cloning of Genes for the 80 and 95 kD Telomerase Subunits

Degenerate primers were designed from peptide sequences with consideration of *Tetrahymena* codon usage frequencies. Martindale, D.W. (1989) *J. Protozol. 36*:29-34. These primers were used in multiple combinations under a variety of PCR conditions. Templates for PCR included *Tetrahymena* macronuclear genomic DNA, or total or poly-A^+ RNA (prepared as described in Ausubel, *et al.*, (1992)

Current Protocols in Molecular Biology, John Wiley & Sons, Inc. and Sambrook, *et al.*, *supra)* from *Tetrahymena* grown and starved as described above). Products from PCR amplification were purified, cloned in *E. coli*, and sequenced by standard protocols. PCR products were confirmed to derive from p80 or p95 gene or cDNA if the PCR product encoded additional peptide sequence not specified by the PCR primer, either as an entirely internal peptide or as sequence adjacent to that specified by the degenerate PCR primer used in the reaction.

PCR products were used to screen an oligo dT-primed *Tetrahymena* cDNA library in λgt10 (Takemasa, *et al.* (1989) *J. Biol. Chem. 264*:19293-19301). Only partial clones (0.8 kb or less) were obtained. Genomic libraries were constructed in Bluescript KS+ (Stratagene) with *Eco*RI or *Cla*I digested *Tetrahymena* genomic DNA. A 3.2 kb clone was obtained that contained most of the gene for p80. A 1.1 kb clone was obtained that contained an internal region of coding sequence for the p95 gene. Fragments containing other portions of the p95 gene were detected by Southern blot of *Eco*RI digested genomic DNA, but were drastically under-represented in the constructed libraries. To obtain the 5′ end of the cDNAs for both genes, a RACE protocol was followed (Gibco # 18374-025) using poly-A+ RNA from starved *Tetrahymena*. To determine the 3′ end of the cDNA for p95, lambda clone sequences were compared with sequence obtained by 3′ RACE. The 3′ RACE was performed based on the protocol above, using oligo dT priming from the mRNA poly-A+ tail for reverse transcription, combined with priming from within the known sequence of the genomic clone for PCR. To determine the 3′ end of the p80 cDNA, the sequences of lambda and genomic clones were compared; the lambda clones obtained for p80 terminate at the 3′ end with poly-A sequence present only as four adenine residues in

the g nomic clone. The results of 3' RACE for p80 support this region as the site of polyadenylation.

Previous determination of the nucleic acid sequence of p95 had indicated the nucleotide at position 405 to be "G" and nucleotides at positions 977-979 to be "CGT". This resulted in an "R" instead of "Q" and "A", respectively, in the encoded protein.

Example 5

Generation of Antibodies to the 80 kD and 95 kD Polypeptides

Synthetic peptides were synthesized that corresponded to two different regions from each protein. Peptides were purchased from Genosys Biotechnologies. These peptides were coupled to Keyhole Lymphet Hemocyanine (KLH) carrier protein via an amino terminal additional cystine residue using standard protocols. Harlow, *et al.* (1988) *Antibodies - A Laboratory Manual*, Cold Spring Harbor Press, Cold Spring Harbor, NY. Each of the peptides coupled to KLH protein were injected into two separate rabbits using standard protocols including periodic boosts with the antigen. Harlow, *et al.*, *supra*. Sera from the rabbits was sampled every several weeks. The animal injections were generated at Hazelton Corporation. Sera from all of the rabbits after 3-4 boosts with the antigen was obtained and initially tested in ELISA assays against the synthetic peptides used to inject each rabbit. Several of the crude sera specifically recognized the peptides. The antibodies were then tested against total *Tetrahymena* and purified telomerase fractions on Western blots. Two rabbits immunized with one peptide had very good titre against the 80 kD protein. The antibodies from each of these rabbits are designated A81 and A82. Similarly, four rabbits immunized with two peptides had a good titre against the 95 kD protein and these antibodies are designated A83, A84,

A85 and A86. A82 had the highest affinity for the 80 kD protein and A86 had the highest affinity for the 95 kD protein.

The antibodies were then affinity purified by binding to a column with the specific peptide coupled to it. Antibodies were eluted from the column first with an acetate buffer (0.1 M NaOAc, pH 4.0) and subsequently with a glycine buffer (0.1 M glycine, pH 2.7) to remove the tighter binding antibodies. The affinity purified antibodies were used for both western blots and immuno-precipitation. Western analysis with sera containing A82 and A86 antibodies showed that both the 80 and 95 kD polypeptides co-purified with telomerase activity throughout the entire column purification scheme (see Example 1). The level of both the 95 and 80 kD protein paralleled the fold increase in enzyme activity at each stage in the purification, consistent with these proteins being telomerase components.

Telomerase activity was specifically immuno-precipitated by the highest affinity antibody directed against the 80 kD protein (A82). Immuno-precipitation was carried out using standard techniques (Harlow, *et al.*, *supra*). The affinity purified antibody was incubated with agarose beads (Pharmacia) coupled to protein G. After the initial binding reaction the highly purified fraction from a non-peak region of the glycerol gradient (see Example 1), was incubated with the beads and telomerase was allowed to bind for 3-4 hours at 4°C. The beads were then spun at a very low speed in an eppendorf tube and the supernatant was removed. The beads were washed three times in T2MG plus 0.1 M KG 0.5% NP-40 and resuspended in the same buffer. Telomerase activity was then assayed in the supernatant, the final wash and pellet fraction for each antibody. Antibody A82 showed telomerase activity in the pellet and

the supernatant was depleted for activity. As a control, an affinity purifi d antibody which did not recognize either the 80 or 95 kD on Western blots (A80) was used in the immuno-precipitation. With this antibody, activity remained in the supernatant as was the case with the lower affinity antibodies directed against the other 80 and 95 kD polypeptides (A83, A84, A85 and A86). These results indicate that the 80 kD polypeptide is a functional component of telomerase.

Example 6

Synthesis of Genes for the 80 and 95 kD Telomerase Subunits

To express the *Tetrahymena* proteins in any organism besides ciliates, the *Tetrahymena* codons that use UAA and UAG to encode glutamine (Martindale, D.W. (1989) *J. Protozol. 36*:29-34) must be replaced. In most eukaryotes these codons denote "stop"; thus, their translation prevents expression of full length proteins. Because there are 44 glutamine codons in the p95 gene and 18 glutamine sites in the p80 gene that require change, these genes were synthesized *de novo* rather than use site-directed mutagenesis to make each substitution. To construct the synthetic genes, it was first established which codon would be used to code for each amino acid. These codons were chosen by their frequency of use in *E. coli* and baculovirus (Rohrmann, G.F. (1986) *J. Gen. Virol. 67*:1499-1513; Zhang, G. *et al.* (1991) *Gene 105*:61-72), two efficient systems for expressing recombinant proteins. The list of codons chosen is shown in Table 2.

TABLE 2

Codons used for Generation of Synthetic Telomerase p80 and p95 Proteins

Amino acids	Codons			
Ala	GCT	GCC	GCA	GCG
Arg	CGT	CGC		
Asn	AAC			
Asp	GAC			
Cys	TGT			
Gln	CAA			
Glu	GAG	GAA		
Gly	GGC			
His	CAC	CAT		
Ile	ATC			
Leu	CTG			
Lys	AAG			
Met	ATG			
Phe	TTC			
Pro	CCG			
Ser	AGC	TCA	TCC	
Thr	ACC	ACT		
Trp	TGG			
Tyr	TAC			
Val	GTT	GTG	GTC	GTA

A GeneWorks (IntelliGenetics) computer program was used to "reverse translate" the protein sequence of both p80 and p95 using the codons in Table 2 for the genetic code. This created a somewhat degenerate DNA sequence due to the degeneracy of the genetic code chosen. The predicted restriction map from this degenerate sequence was

then examined to find unique restriction sites approximately every 300 bp in each gene sequence. Where necessary, restriction sites were eliminated individually by choosing a different codon to encode a particular amino acid. This process was re-iterated until a unique DNA sequence was obtained that had the appropriate placement of restriction sites. The final DNA sequence including engineered restriction sites at the 5' and 3' end used for cloning are shown in Figures 10 and 11.

To synthesize the two genes shown in Figures 10 and 11, a set of 28 overlapping oligonucleotides were synthesized for p80 (primer sets 3 and 4, Figures 7 and 8, respectively) and a set of 34 were synthesized for p95 (primer sets 1, 2 and 3, Figures 5, 6 and 7, respectively) and the genes were constructed by overlap extension PCR. Prodromou, C. and L.H. Pearl (1992) *Protein Engineering* 5:827-829; Bambot, S.B. and A.J. Russell (1993) *PCR Meths. and Apps.* 2:266-271. The oligonucleotides were purchased from Bioserve Biotechnologies (Laurel, MD). Each oligonucleotide is approximately 100 nt long and is designed to overlap with its compliment to give a hybrid of 20 base pairs. Each oligonucleotide also has a phosphorothioate linkage in place of the usual phosphodiester at the 3' end of the oligonucleotide. This phosphorothioate will prevent exonuclease removal of the 20 bp hybrid overlap during the initial polymerase elongation step. Skerr, A. (1992) *Nucl. Acid Res.* 20:3551-3554.

The p80 gene was constructed in two pieces by combining the first set of oligonucleotides (primer set 4, Figures 8A-8B) pair-wise, then using PCR to amplify the entire region as described. The second half was constructed in a similar manner using primer set 5 (Figure 9). Each half of the gene was cloned into the plasmid Bluescript and sequenced in its entirety to be sure no new mutations were introduced. The 5' half was cloned on a *Bam*

HI-*Eco*RI fragment and the 3' half was cloned on a *Eco*RI-*Kpn*I fragment. The full-length gene was then constructed by combining the two fragments in pBluescript (Stratagene) in the appropriate order.

The p95 gene was constructed in three pieces by combining each set of oligonucleotides pair-wise and then using PCR to amplify the entire region. The first fragment was constructed using primer set 1, the second fragment was constructed with primer set 2 and the third with primer set 3. Each of the three fragments of the gene were cloned into the plasmid pSE280 (Invitrogen) and sequenced in its entirety. The 5' fragment (Fragment and primer set 1) was cloned on a *Nco*I-*Bst*BI fragment, the internal piece (Fragment and primer set 2) was on a *Bst*BI-*Eco*RI fragment, and the 3' fragment (Fragment and primer set 3) was cloned on an *Eco*RI-*Hind*III fragment. The full-length gene was then constructed by first combining fragments 1 and 2 in the pSE280 plasmid, and subsequently adding the 3' fragment to complete the gene.

Example 7

Expression and Purification of Recombinant p80 and p95 Proteins from Cells

The p80 and p95 proteins is expressed in *E. coli* and baculovirus by cloning the full length construct into pRSET and pBlueBac vectors respectively (Invitrogen) by methods known to those of skill in the art (See, *e.g.*, Ausubel, *supra*; Sambrook, *supra*). These vectors allow expression and purification of the recombinant proteins. The p80 is cloned into the *Bam*HI and *Hind*III sites of pRSET and pBlueBac and the p95 is cloned into the *Nco*I and *Hind*III sites. Transcription and translation of these constructs generates the recombinant proteins followed by a series of 6 histidines (His-tag), separated by an EK (Enterokinase) protease cleavage site. This allows each protein to be

purified by processing over a Ni^{++} chelating column. The purified protein is removed from the His tag by digestion with the protease Enterokinase.

Example 8

Cloning of Human Telomerase Protein Components

Two general approaches can be taken to cloning the human genes: DNA sequence based approaches and antibody directed approaches. The first approach takes advantage of the DNA sequence of the *Tetrahymena* genes to directly identify the human homologues. Those skilled in the art will recognize three different strategies that are used to clone homologues based on DNA sequence: (1) direct hybridization of human genomic or cDNA libraries with the *Tetrahymena* gene; (2) identification of conserved regions in telomerase protein in other species and PCR amplification of a human gene based on these regions; and (3) systematic strategy to saturate all regions of the telomerase genes with PCR probes and identification of a human homologue using PCR to "walk" along the length of the gene. All of the methodology is based on standard molecular genetic laboratory procedures (Sambrook, *et al.*, *supra*).

In the first strategy, the *Tetrahymena* gene is used to probe human genomic DNA and mRNA blots at a series of increasing stringencies. When specific bands are identified, the cDNA or genomic library can be probed at a similar stringency to identify the gene for the human homologue. Positive phage is then restriction mapped, subcloned, and sequenced.

The second strategy involves cloning the telomerase proteins from other ciliates first because telomerase proteins may have only limited conservation at the DNA sequence level between humans and *Tetrahymena*. Then the

mammalian counterparts are cloned using information obtained from these cDNAs. Thus, using the cloned *Tetrahymena* genes, libraries from distantly related *Tetrahymena* or *Oxytricha* and *Euplotes* can be probed at medium stringency to identify genes which cross hybridize. Since the ciliates *Oxytricha* and *Euplotes* have telomerase enzymes which are functionally similar to the *Tetrahymena* telomerase (Lingner, *et al.* (1994) *Genes Dev. 8*:1984-1998; Shippen-Lentz, D. and E.H. Blackburn (1990) *Science 247*: 546-552), it is likely that homologue proteins can be identified with this method. The genes for both the p95 and p80 homologues from both ciliates can be fully sequenced and regions of the highest degree of similarity between the different species can be identified. Three conserved regions are chosen for each protein to use in Reverse transcriptase PCR-based approaches to cloning the human gene. The same approach is taken to clone the genes for both the p80 and p95 genes. Degenerate oligonucleotides encoding the conserved regions in the *Tetrahymena, Oxytricha* and *Euplotes* telomerase proteins are synthesized using human translational codon biases. PCR is initially carried out with two of the three oligonucleotides. The 3' most oligonucleotide (oligo 1) can be complementary to the mRNA. Thus cDNA is synthesized from isolated mRNA using oligo 1 as a primer. The 5' most oligo (oligo 2) can be oriented 5' to 3' in the direction opposite to oligo 1 and can be identical in sequence to the mRNA strand. This oligonucleotide is then used in a PCR step along with oligo 1 to amplify the region between oligo 1 and oligo 2. Finally a third primer (oligo 3) is directed against a conserved region of the protein which lies between the regions targeted by oligos 1 and 2. The sequence of the oligo will be complementary to the mRNA. The PCR product amplified with oligos 1 and 2 is reamplified using oligo 2 and oligo 3. This is to assure

that the specific products generated all have three conserved regions of the telomerase proteins. The PCR products are sequenced to identify those that contain protein homologues.

The third strategy is a systematic scanning approach to find regions of homology between the *Tetrahymena* and human genes, and then to amplify these regions by PCR. Using the protein sequence of the *Tetrahymena* genes as a guide, a series of primer oligonucleotides is generated that encode regions of the *Tetrahymena* genes yet utilize the human codon bias in the DNA sequence. Initially, a set of primers differing in the region to which they hybridize by 10 amino acids, is generated against the 3' end of the gene. These are used in an RT PCR reaction to generate cDNA. Next, a set of primers oriented from the 5' of the gene toward the 3' end are used to amplify the cDNA. All possible combinations of two PCR primers from the 5' end and 3' end can be used together to identify bands that are the size expected for the regions in the *Tetrahymena* protein. If specific products are generated they are reamplified using primers that should anneal within the initial two primers. Specific products which are not generated by any primer alone are subcloned and sequenced.

Example 9

Antibody directed approaches to cloning human telomerase homologues

This method takes advantage of conserved epitopes on the protein surface that are recognized by antibodies. A series of antibodies are made to various regions of both the p80 and p95 proteins. Antibodies such as those described in Example 5 can be used. The antigens for antibody production are generated as synthetic peptides or as fusion proteins. It is faster to produce synthetic

peptides, since the fusion proteins do not have to be generated first; however, these peptides may not give as high a titer compared to synthetic peptides. Approximately 10-15 residues of the peptides are selected, preferably at the 5′ and 3′ ends of the protein since these residues are likely to be unstructured and thus antigenic. In addition, computer analysis can be applied to determine the hydrophilicity and predicted secondary structure of the protein to choose regions which are likely to be unstructured and near the surface of the protein. The synthetic peptide is coupled to a carrier such as KLH and used to inoculate rabbits. If the appropriate amino acids for coupling to the carrier are not present in the peptide, a linker cysteine residue is added to the N-terminus.

Fusion proteins are generated with the T7 polymerase system (Studier, *et al.* (1990) *Meth. Enzymol*, *185*:60-89) and purified for inoculation into rabbits or mice. The sera is then screened on Western blots using extracts from *E. coli* expressing the cloned protein or purified fractions. The positive antibodies are tested for their ability to recognize the 95 or 80 kD proteins on Western blots and/or to specifically immunoprecipitate telomerase RNA or otherwise inhibit telomerase activity. As a control, the ability of the anti-peptide antibody to precipitate telomerase RNA should be abolished when it is pre-incubated with the peptide. Because mouse and rabbit sera and monoclonal culture medium inhibit telomerase activity (L. Harrington, unpublished results), affinity purified IgG is used to test the ability of the antibodies to inhibit telomerase activity.

Both polyclonal and monoclonal antibodies against *Tetrahymena* telomerase proteins can be used to identify cross-reacting telomerase proteins from human and mouse cells on Western blots. If a positive signal is found, purified fractions of human telomerase is used to determine

if the reactive band co-purifies with telomerase activity. Evidence of co-purification indicates that the cross-reacting band is a component of human telomerase. Antibodies which give the best signal are then used to probe expression libraries of lambda GT11. If monoclonal antibodies are used, several different antibodies are pooled for probing the expression libraries. For polyclonal antibodies, two or three different antibodies are used on duplicate plates. Only those phage which light up with both probes are considered positive. These plaques are purified and the inserts subcloned and sequenced.

Example 10

Cloning and use of the mouse homologue

The same two procedures of DNA homology or antibody cross-reactivity describe above can also be used in parallel to identify the mouse telomerase protein components. Mouse telomerase clones can be useful in testing cancer therapies and for understanding the biology of mammalian telomerase. Identification of mouse telomerase will also allow the use of transgenic mice to test the roles of telomere length and telomerase *in vivo*. Once either the mouse or human homologue has been identified, either clone is applied to deduce the sequence of the other organism. Sequence similarity is high between human and mouse genes making it a straightforward process to obtain the clone for one with a probe from the other. Both genomic and cDNA libraries are then plated and probed at a moderate stringency to identify cross hybridizing plaques. The positive plaques are then selected, restriction mapped and sequenced, to determine if the telomerase protein homologue has been cloned. Further functional analysis, such as reconstitution and gene disruption is then applied with the human and mouse clones.

Equivalents

Those skilled in the art will recognize, or be able to ascertain using no more than routine experimentation, many equivalents to the specific embodiments of the invention described specifically herein. Such equivalents are intended to be encompassed in the scope of the following claims.

Claims

We claim:

1. A substantially pure telomerase protein component.

2. A substantially pure protein component of Claim 1, which is a *Tetrahymena* telomerase protein component.

3. Isolated DNA which encodes a telomerase protein component and is identical to or substantially homologous to the nucleotide sequence of SEQ ID NO:1 or SEQ ID NO:3.

4. DNA which hybridizes under moderate stringency conditions to the DNA according to Claim 3.

5. Isolated RNA transcribed from or complementary to the DNA of Claim 3.

6. Isolated RNA transcribed from or complementary to the DNA of Claim 4.

7. A polypeptide encoded by DNA of Claim 3.

8. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO:5.

9. An isolated polypeptide comprising the amino acid sequence of SEQ ID NO:7.

10. Isolated DNA which encodes a polypeptide identical to or substantially equivalent to the amino acid sequence of SEQ ID NO:2 or SEQ ID NO:4.

11 An anti-telomerase antibody which binds a peptide comprising an amino acid sequence selected from the group consisting of: AEGYSDINVRG, QNEFQFNNVK and EFGLEPNILT.

12. An anti-telomerase antibody which binds all or a portion of a substantially pure telomerase protein component.

13. A method of detecting the presence of immortal cells or a predisposition to immortalization of cells in a eukaryotic tissue sample or a sample of eukaryotic cells, comprising the steps of:

a) obtaining a tissue sample or a sample of cells from the eukaryote; and

b) determining the presence of telomerase in the sample, wherein if the sample demonstrates presence of telomerase, immortal cells or the predisposition to immortalization is present.

14. A method of detecting a disease caused by a eukaryotic microbial organism in a eukaryotic tissue sample or a sample of eukaryotic cells, comprising the steps of:

a) obtaining a tissue sample or a sample of cells from the eukaryote; and

b) determining the telomerase in the sample, wherein if the sample demonstrates telomerase of a eukaryotic microbe, a disease caused by a eukaryotic microbial organism is present.

15. A method of detecting telomerase in a eukaryotic tissue sample or a sample of eukaryotic cells, comprising the steps of:

a) obtaining a tissue sample or a sample of cells from the eukaryote;

b) treating the sample to render telomerase available for binding by anti-telomerase antibodies, thereby producing a treated sample;

c) contacting the treated sample with anti-telomerase antibodies; and

d) detecting binding of the antibodies to telomerase, wherein if binding occurs, telomerase is present.

16. The method of Claim 15, wherein the presence of telomerase indicates a predisposition to cancer or the presence of cancer.

17. The method of Claim 15, wherein the presence of telomerase indicates the presence of immortal cells or a predisposition to immortalization.

18. The method of Claim 15, wherein the presence of telomerase of a eukaryotic microbe indicates the presence of a disease caused by a eukaryotic microbial organism.

19. A method of identifying a compound that inhibits, destroys, or interferes with telomerase activity in eukaryotic cells, comprising administering the compound to a *Tetrahymena* cell and measuring activity of *Tetrahymena* telomerase in the cell, wherein if the *Tetrahymena* telomerase activity is reduced, the compound is a telomerase inhibitor.

20. A method of identifying a compound that inhibits, destroys, or interferes with telomerase activity in mammalian, including human, cells, comprising administering the compound to eukaryotic cells expressing telomerase activity and measuring the

activity of telomerase, wherein if the telomerase activity is reduced, said compound or compounds are identified as telomerase inhibitors.

21. A therapeutic or diagnostic compound comprising an inhibitor of telomerase activity in combination with a pharmaceutically acceptable carrier, diluent or excipient.

22. A therapeutic or diagnostic compound comprising an amino acid sequence encoded by isolated DNA according to Claim 10 in combination with a pharmaceutically acceptable carrier, diluent, or excipient.

23. A pharmaceutical composition comprising all or substantially all of one or both polypeptides according to SEQ ID NO:2 or SEQ ID NO:4, in combination with a pharmaceutically acceptable diluent, excipient, or carrier.

24. A process for the preparation of a therapeutic or diagnostic composition comprising combining an anti-telomerase compound together with a pharmaceutically acceptable excipient, diluent, or carrier.

25. A method for treating a disease caused by a eukaryotic microorganism in a mammal comprising administering to the mammal an amount of a telomerase inhibitor effective to inhibit telomerase activity in the microorganism.

26. A method of inhibiting the activity of eukaryotic microbial parasites, especially fungal and protozoan parasites, in mammals comprising administering to the mammal an amount of a telomerase inhibitor effective

to inhibit the activity of the eukaryotic microbial parasites.

27. A method of therapeutic treatment of a human or animal suffering with a disorder associated with an abnormal level of telomerase activity comprising inhibiting the production of telomerase if said level is too high or administering telomerase if said level is too low.

28. A transgenic eukaryotic cell or organism containing the DNA sequence of Claim 3 or a sequence complementary to said sequence.

29. A transgenic prokaryotic cell containing the DNA sequence of Claim 3 or a sequence complementary to said sequence.

30. A transgenic eukaryotic cell or organism containing the nucleotide sequences SEQ ID NO:1 and SEQ ID NO:3.

31. A process for producing recombinant telomerase, comprising the steps of:

(a) producing an expression vector which includes DNA which encodes a telomerase molecule;

(b) transfecting or infecting a host cell with the vector; and

(c) culturing the transfected or infected cell line to produce the encoded telomerase.

32. A process for producing a recombinant anti-telomerase antibody comprising:

(a) producing an expression vector which includes DNA which encodes an anti-telomerase antibody;

(b) transfecting or infecting a host cell with the vector, thereby producing a transfected or infected host cell; and

(c) culturing the transfected or infected cell to produce the anti-telomerase antibody.

33. A DNA sequence comprising SEQ ID NO:8 or SEQ ID NO:9.

34. A synthetic telomerase protein component encoded by a DNA sequence according to Claim 33.

35. A DNA or RNA sequence that hybridizes to a DNA sequence according to Claim 33.

36. An expression vector comprising DNA selected from the group consisting of any of the primers 1-34, F1-F14, and R1-R14.

37. A host cell comprising an expression vector comprising DNA encoding a telomerase protein component or a fragment thereof.

38. A method for producing a telomerase protein component or a fragment thereof, comprising the step of culturing a host cell of Claim 37 under conditions which permit production of the telomerase protein component or fragment thereof.

39. A method according to Claim 38 wherein two or more telomerase protein components or fragments are produced in the same cell.

40. The method of Claim 38, further comprising the step of purifying the telomerase protein component or fragment.

```
   1 aactcattta attactaatt taatcaacaa gattgataaa aagcagtaaa taaaacccaa
  61 tagatttaat ttagaaagta tcaattgaaa aatggaaatt gaaaacaact aagcacaata
 121 gccaaaagcc gaaaaattgt ggtgggaact tgaattagag atgcaagaaa accaaaatga
 181 tatataagtt agggttaaga ttgacgatcc taagcaatat ctcgtgaacg tcactgcagc
 241 atgtttgttg taggaaggta gttactacta agataaagat gaaagaagat atatcatcac
 301 taaagcactt cttgaggtgg ctgagtctga tcctgagttc atctgctagt tggcagtcta
 361 catccgtaat gaactttaca tcagaactac cactaactac attgtagcat tttgtgttgt
 421 ccacaagaat actcaaccat tcatcgaaaa gtacttcaac aaagcagtac ttttgcctaa
 481 tgacttactg gaagtctgtg aatttgcata ggttctctat atttttgatg caactgaatt
 541 caaaaatttg tatcttgata ggatactttc ataagatatt cgtaaggaac tcactttccg
 601 taagtgttta caaagatgcg tcagaagcaa gttttctgaa ttcaacgaat actaacttgg
 661 taagtattgc actgaatcct aacgtaagaa aacaatgttc cgttacctct cagttaccaa
 721 caagtaaaag tgggattaaa ctaagaagaa gagaaaagag aatctcttaa ccaaacttta
 781 ggcaataaag gaatctgaag ataagtccaa gagagaaact ggagacataa tgaacgttga
 841 agatgcaatc aaggctttaa aaccagcagt tatgaagaaa atagccaaga gatagaatgc
 901 catgaagaaa cacatgaagg cacctaaaat tcctaactct accttggaat caaagtactt
 961 gaccttcaag gatctcatta agttctgcca tatttctgag cctaaagaaa gagtctataa
1021 gatccttggt aaaaaatacc ctaagaccga agaggaatac aaagcagcct ttggtgattc
1081 tgcatctgca cccttcaatc ctgaattggc tggaaagcgt atgaagattg aaatctctaa
1141 aacatgggaa aatgaactca gtgcaaaagg caacactgct gaggtttggg ataatttaat
1201 ttcaagcaat taactcccat atatggccat gttacgtaac ttgtctaaca tcttaaaagc
1261 cggtgtttca gatactacac actctattgt gatcaacaag atttgtgagc ccaaggccgt
1321 tgagaactcc aagatgttcc ctcttcaatt ctttagtgcc attgaagctg ttaatgaagc
1381 agttactaag ggattcaagg ccaagaagag agaaaatatg aatcttaaag gtcaaatcga
1441 agcagtaaag gaagttgttg aaaaaaccga tgaagagaag aaagatatgg agttggagta
1501 aaccgaagaa ggagaatttg ttaaagtcaa cgaaggaatt ggcaagcaat acattaactc
1561 cattgaactt gcaatcaaga tagcagttaa caagaattta gatgaaatca aaggacacac
1621 tgcaatcttc tctgatgttt ctggttctat gagtacctca atgtcaggtg gagccaagaa
1681 gtatggttcc gttcgtactt gtctcgagtg tgcattagtc cttggtttga tggtaaaata
1741 acgttgtgaa aagtcctcat tctacatctt cagttcacct agttctcaat gcaataagtg
1801 ttacttagaa gttgatctcc ctggagacga actccgtcct tctatgtaaa aacttttgca
1861 agagaaagga aaacttggtg gtggtactga tttcccctat gagtgcattg atgaatggac
1921 aaagaataaa actcacgtag acaatatcgt tattttgtct gatatgatga ttgcagaagg
1981 atattcagat atcaatgtta gaggcagttc cattgttaac agcatcaaaa agtacaagga
2041 tgaagtaaat cctaacatta aaatctttgc agttgactta gaaggttacg gaaagtgcct
2101 taatctaggt gatgagttca atgaaaacaa ctacatcaag atattcggta tgagcgattc
2161 aatcttaaag ttcatttcag ccaagcaagg aggagcaaat atggtcgaag ttatcaaaaa
2221 ctttgccctt caaaaaatag gacaaaagtg agtttcttga gattcttcta taacaaaaat
2281 ctcaccccac ttttttgttt tattgcatag ccattatgaa atttaaatta ttatctattt
2341 atttaagtta cttacatagt ttatgtatcg cagtctatta gcctattcaa atgattctgc
2401 aaagaacaaa aaagattaaa a
```

FIGURE 1

MEIENNQAQQPKAEKLWWELELEMQENQNDIQVRVKIDDPKQYL
VNVTAACLLQEGSYYQDKDERRYIITKALLEVAESDPEFICQLAVYIRNELYIRTTTN
YIVAFCVVHKNTQPFIEKYFNKAVLLPNDLLEVCEFAQVLYIFDATEFKNLYLDRILS
QDIRKELTFRKCLQRCVRSKFSEFNEYQLGKYCTESQRKKTMFRYLSVTNKQKWDQTK
KKRKENLLTKLQAIKESEDKSKRETGDIMNVEDAIKALKPAVMKKIAKRQNAMKKHMK
APKIPNSTLESKYLTFKDLIKFCHISEPKERVYKILGKKYPKTEEEYKAAFGDSASAP
FNPELAGKRMKIEISKTWENELSAKGNTAEVWDNLISSNQLPYMAMLRNLSNILKAGV
SDTTHSIVINKICEPKAVENSKMFPLQFFSAIEAVNEAVTKGFKAKKRENMNLKGQIE
AVKEVVEKTDEEKKDMELEQTEEGEFVKVNEGIGKQYINSIELAIKIAVNKNLDEIKG
HTAIFSDVSGSMSTSMSGGAKKYGSVRTCLECALVLGLMVKQRCEKSSFYIFSSPSSQ
CNKCYLEVDLPGDELRPSMQKLLQEKGKLGGGTDFPYECIDEWTKNKTHVDNIVILSD
MMIAEGYSDINVRGSSIVNSIKKYKDEVNPNIKIFAVDLEGYGKCLNLGDEFNENNYI
KIFGMSDSILKFISAKQGGANMVEVIKNFALQKIGQK

FIGURE 2

```
   1 tcaatactat taattaataa ataaaaaaaa gcaaactaca aagaaaatgt caaggcgtaa
  61 ctaaaaaaag ccataggctc ctataggcaa tgaaacaaat cttgattttg tattacaaaa
 121 tctagaagtt tacaaaagcc agattgagca ttataagacc tagtagtaat agatcaaaga
 181 ggaggatctc aagcttttaa agttcaaaaa ttaagattag gatggaaact ctggcaacga
 241 tgatgatgat gaagaaaaca actcaaataa ataataagaa ttattaagga gagtcaatta
 301 gattaagtag caagtttaat tgataaaaaa agttggttct aaggtagaga aagatttgaa
 361 tttgaacgaa gatgaaaaca aaaagaatgg actttctgaa tagcaagtga aagaagagta
 421 attaagaacg attactgaag aataggttaa gtattaaaat ttagtattta acatggacta
 481 ccagttagat ttaaatgaga gtggtggcca tagaagacac agaagagaaa cagattatga
 541 tactgaaaaa tggtttgaaa tatctcatga ccaaaaaaat tatgtatcaa tttacgccaa
 601 ctaaaagaca tcatattgtt ggtggcttaa agattatttt aataaaaaca attatgatca
 661 tcttaatgta agcattaaca gactagaaac tgaagccgaa ttctatgcct ttgatgattt
 721 ttcacaaaca atcaaactta ctaataattc ttactagact gttaacatag acgttaattt
 781 tgataataat ctctgtatac tcgcattgct tagattttta ttatcactag aaagattcaa
 841 tattttgaat ataagatctt cttatacaag aaattaatat aattttgaga aaattggtga
 901 gctacttgaa actatcttcg cagttgtctt ttctcatcgc cacttacaag gcattcattt
 961 acaagttcct tgcgaagcgt tctaatattt agttaactcc tcatcataaa ttagcgttaa
1021 agatagctaa ttataggtat actctttctc tacagactta aaattagttg acactaacaa
1081 agtccaagat tattttaagt tcttataaga attcgctcgt ttgactcatg taagctagta
1141 ggctatccca gttagtgcta ctaacgctgt agagaacctc aatgttttac ttaaaaaggt
1201 caagcatgct aatcttaatt tagtttctat ccctacctaa ttcaattttg atttctactt
1261 tgttaattta taacatttga aattagagtt tggattagaa ccaaatattt tgacaaaaca
1321 aaagcttgaa aatctacttt tgagtataaa ataatcaaaa aatcttaaat ttttaagatt
1381 aaacttttac acctacgttg cttaagaaac ctccagaaaa cagatattaa aacaagctac
1441 aacaatcaaa aatctcaaaa acaataaaaa tcaagaagaa actcctgaaa ctaaagatga
1501 aactccaagc gaaagcacaa gtggtatgaa atttttgat catctttctg aattaaccga
1561 gcttgaagat ttcagcgtta acttgtaagc tacccaagaa atttatgata gcttgcacaa
1621 acttttgatt agatcaacaa atttaaagaa gttcaaatta agttacaaat atgaaatgga
1681 aaagagtaaa atggatacat tcatagatct taagaatatt tatgaaacct taaacaatct
1741 taaaagatgc tctgttaata tatcaaatcc tcatggaaac atttcttatg aactgacaaa
1801 taaagattct acttttata aatttaagct gaccttaaac taagaattat aacacgctaa
1861 gtatactttt aagtagaacg aattttaatt taataacgtt aaaagtgcaa aaattgaatc
1921 ttcctcatta gaaagcttag aagatattga tagtctttgc aaatctattg cttcttgtaa
1981 aaatttacaa aatgttaata ttatcgccag tttgctctat cccaacaata tttagaaaaa
2041 tcctttcaat aagcccaatc ttctattttt caagcaattt gaataattga aaaatttgga
2101 aaatgtatct atcaactgta ttcttgatca gcatatactt aattctattt cagaattctt
2161 agaaaagaat aaaaaaataa aagcattcat tttgaaaaga tattatttat tacaatatta
2221 tcttgattat actaaattat ttaaaacact tcaatagtta cctgaattaa attaagttta
2281 cattaattag caattagaag aattgactgt gagtgaagta cataagtaag tatgggaaaa
2341 ccacaagcaa aaagctttct atgaaccatt atgtgagttt atcaaagaat catcctaaac
2401 cctttagcta atagattttg accaaaacac tgtaagtgat gactctatta aaaagatttt
2461 agaatctata tctgagtcta agtatcatca ttatttgaga ttgaacccta gttaatctag
2521 cagtttaatt aaatctgaaa acgaagaaat ttaagaactt ctcaaagctt gcgacgaaaa
2581 aggtgtttta gtaaaagcat actataaatt ccctctatgt ttaccaactg gtacttatta
2641 cgattacaat tcagatagat ggtgattaat taaatattag tttaaataaa tattaaatat
2701 tgaatatttc tttgcttatt atttgaataa tacatacaat agtcattttt agtgttttga
2761 atatatttta gttatttaat tcattatttt aagtaaataa ttatttttca atcatttttt
2821 aaaaaatcg
```

FIGURE 3

MSRRNQKKPQAPIGNETNLDFVLQNLEVYKSQIEHYKTQQQQIK
EEDLKLLKFKNQDQDGNSGNDDDDEENNSNKQQELLRRVNQIKQQVQLIKKVGSKVEK
DLNLNEDENKKNGLSEQQVKEEQLRTITEEQVKYQNLVFNMDYQLDLNESGGHRRHRR
ETDYDTEKWFEISHDQKNYVSIYANQKTSYCWWLKDYFNKNNYDHLNVSINRLETEAE
FYAFDDFSQTIKLTNNSYQTVNIDVNFDNNLCILALLRFLLSLERFNILNIRSSYTRN
QYNFEKIGELLETIFAVVFSHRHLQGIHLQVPCEAFQYLVNSSSQISVKDSQLQVYSF
STDLKLVDTNKVQDYFKFLQEFPRLTHVSQQAIPVSATNAVENLNVLLKKVKHANLNL
VSIPTQFNFDFYFVNLQHLKLEFGLEPNILTKQKLENLLLSIKQSKNLKFLRLNFYTY
VAQETSRKQILKQATTIKNLKNNKNQEETPETKDETPSESTSGMKFFDHLSELTELED
FSVNLQATQEIYDSLHKLLIRSTNLKKFKLSYKYEMEKSKMDTFIDLKNIYETLNNLK
RCSVNISNPHGNISYELTNKDSTFYKFKLTLNQELQHAKYTFKQNEFQFNNVKSAKIE
SSSLESLEDIDSLCKSIASCKNLQNVNIIASLLYPNNIQKNPFNKPNLLFFKQFEQLK
NLENVSINCILDQHILNSISEFLEKNKKIKAFILKRYYLLQYYLDYTKLFKTLQQLPE
LNQVYINQQLEELTVSEVHKQVWENHKQKAFYEPLCEFIKESSQTLQLIDFDQNTVSD
DSIKKILESISESKYHHYLRLNPSQSSSLIKSENEEIQELLKACDEKGVLVKAYYKFP
LCLPTGTYYDYNSDRW

FIGURE 4

PRIMER SET 1

```
primer1 Length: 100
   1  GGGGCCATGG ATGAGCCGTC GTAACCAAAA GAAGCCGCAA GCTCCGATCG
  51  GCAACGAGAC CAACCTGGAC TTCGTTCTGC AAAACCTGGA GGTTTACAAG

primer2  Length: 100  ..
   1  CTTGGTTCTT GAACTTCAGC AGCTTCAGGT CCTCCTCCTT GATTTGTTGT
  51  TGTTGGGTCT TGTAGTGCTC GATTTGGCTC TTGTAAACCT CCAGGTTTTG

primer3 Length: 99
   1  TGCTGAAGTT CAAGAACCAA GACCAAGACG GCAACAGCGG CAACGACGAC
  51  GACGACGAGG AGAACAACAG CAACAAGCAA CAAGAGCTGC TGCGTCGTG

primer4  Length: 100
   1  CGTCCTCGTT CAGGTTCAGG TCCTTCTCAA CCTTGCTGCC AACCTTCTTG
  51  ATCAGTTGAA CTTGTTGCTT GATTTGGTTA ACACGACGCA GCAGCTCTTG

primer5  Length: 100
   1  CCTGAACCTG AACGAGGACG AGAACAAGAA GAACGGCCTG AGCGAGCAAC
  51  AAGTTAAGGA GGAGCAACTG CGTACCATCA CCGAGGAGCA AGTTAAGTAC

primer6 Length: 100
   1  TCGGTCTCGC GACGGTGACG ACGGTGGCCG CCGCTCTCGT TCAGGTCCAG
  51  TTGGTAGTCC ATGTTGAAAA CCAGGTTTTG GTACTTAACT TGCTCCTCGG

primer7  Length: 99 .
   1  CGTCACCGTC GCGAGACCGA CTACGACACC GAGAAGTGGT TCGAGATCAG
  51  CCACGACCAA AAGAACTACG TTAGCATCTA CGCTAACCAA AAGACCAGC

primer8  Length: 100
   1  CTCGGTCTCC AGACGGTTGA TGCTAACGTT CAGGTGGTCG TAGTTGTTCT
  51  TGTTGAAGTA GTCCTTCAGC CACCAACAGT AGCTGGTCTT TTGGTTAGCG

primer9 Length: 99
   1  TCAACCGTCT GGAGACCGAG GCTGAGTTCT ACGCTTTCGA CGACTTCAGC
  51  CAAACCATCA AGCTGACCAA CAACAGCTAC CAAACCGTTA ACATCGACG

primer10  Length: 100
   1  GATGTTCAGG ATGTTGAAAC GCTCCAGGCT CAGCAGGAAA CGCAGCAGAG
  51  CCAGGATACA CAGGTTGTTG TCGAAGTTGA CGTCGATGTT AACGGTTTGG

primer11 Length: 99
   1  CGTTTCAACA TCCTGAACAT CCGTAGCAGC TACACCCGTA ACCAATACAA
  51  CTTCGAAAAG ATCGGCGAGC TGCTGGAGAC CATCTTCGCT GTTGTTTTC

primer12   Length: 100
   1  TTGGCTGCTG CTGTTAACCA GGTATTGGAA AGCCTCACAC GGAACTTGCA
  51  GGTGGATGCC TTGCAGGTGA CGGTGGCTGA AAACAACAGC GAAGATGGTC
```

FIGURE 5

PRIMER SET 2

primer13 Length: 99
1 TGGTTAACAG CAGCAGCCAA ATCAGCGTTA AGGACAGCCA ACTGCAAGTT
51 TACAGCTTCA GCACCGACCT GAAGCTGGTT GACACCAACA AGGTTCAAG

primer14 Length: 100
1 CGTTGGTAGC GCTAACCGGG ATAGCTTGTT GGCTCACGTG GGTCAGACGC
51 GGGAACTCTT GCAGGAACTT GAAGTAGTCT TGAACCTTGT TGGTGTCAAC

primer15 Length: 99
1 CCGGTTAGCG CTACCAACGC TGTTGAGAAC CTGAACGTTC TGCTGAAGAA
51 GGTTAAGCAC GCTAACCTGA ACCTGGTTAG CATCCCGACC CAATTCAAC

primer16 Length: 100 ..
1 AGCTTTTGCT TGGTCAGGAT GTTCGGCTCC AGGCCGAACT CCAGCTTCAG
51 GTGTTGCAGG TTAACGAAGT AGAAGTCGAA GTTGAATTGG GTCGGGATGC

primer17 Length: 98
1 CATCCTGACC AAGCAAAAGC TGGAGAACCT GCTGCTGAGC ATCAAGCAAA
51 GCAAGAACCT GAAGTTCCTG CGTCTGAACT TCTACACCTA CGTTGCTC

primer18 Length: 100
1 AGTCTCCTCT TGGTTCTTGT TGTTCTTCAG GTTCTTGATG GTGGTAGCTT
51 GCTTCAGGAT TTGCTTACGG CTGGTCTCTT GAGCAACGTA GGTGTAGAAG

primer19 Length: 100
1 AACAAGAACC AAGAGGAGAC TCCGGAGACC AAGGACGAGA CCCCGAGCGA
51 GAGCACCAGC GGCATGAAGT TCTTCGACCA CCTGAGCGAG CTGACCGAGC

primer20 Length: 100..
1 GGTTGGTGCT ACGGATCAGC AGCTTGTGCA GGCTGTCGTA GATCTCTTGG
51 GTAGCTTGCA GGTTAACGCT GAAGTCCTCC AGCTCGGTCA GCTCGCTCAG

primer21 Length: 100
1 GCTGATCCGT AGCACCAACC TGAAGAAGTT CAAGCTGAGC TACAAGTACG
51 AGATGGAGAA GAGCAAGATG GACACCTTCA TCGATCTGAA GAACATCTAC

primer22 Length: 100
1 GTTGGTCAGC TCGTAGCTGA TGTTGCCGTG CGGGTTGCTG ATGTTAACGC
51 TACAACGCTT CAGGTTGTTC AGGGTCTCGT AGATGTTCTT CAGATCGATG

primer23 Length: 99
1 TCAGCTACGA GCTGACCAAC AAGGACAGCA CCTTCTACAA GTTCAAGCTG
51 ACCCTGAACC AAGAGCTGCA ACACGCTAAG TACACCTTCA AGCAAAACG

primer24 Length: 100
1 ACACAGGCTG TCGATGTCCT CCAGGCTCTC CAGGCTGCTG CTCTCGATCT
51 TAGCGCTCTT AACGTTGTTG AATTGGAATT CGTTTTGCTT GAAGGTGTAC

FIGURE 6

PRIMER SET 3

```
primer25  Length: 101
   1  AGGACATCGA CAGCCTGTGT AAGAGCATCG CCAGCTGTAA GAACCTGCAA
  51  AACGTTAACA TCATCGCTAG CCTGCTGTAC CCGAACAACA TCCAAAAGAA
 101  C

primer26 Length: 100
   1  TACAGTTGAT GCTAACGTTC TCCAGGTTCT TCAGTTGCTC GAATTGCTTG
  51  AAGAACAGCA GGTTCGGCTT GTTGAACGGG TTCTTTTGGA TGTTGTTCGG

primer27 Length: 96
   1  GAGAACGTTA GCATCAACTG TATCCTGGAC CAACACATCC TGAACAGCAT
  51  CAGCGAGTTC CTGGAGAAGA ACAAGAAGAT CAAGGCTTTC ATCCTG

primer28 Length: 100
   1  GTTCAGCTCC GGCAGTTGTT GCAGGGTCTT GAACAGCTTG GTGTAGTCCA
  51  GGTAGTATTG CAGCAGGTAG TAACGCTTCA GGATGAAAGC CTTGATCTTC

primer29 Length: 101
   1  AACAACTGCC GGAGCTGAAC CAAGTTTACA TCAACCAACA ACTGGAGGAG
  51  CTGACCGTTA GCGAGGTTCA CAAGCAAGTT TGGGAGAACC ACAAGCAAAA
 101  G

primer30 Length: 100  .
   1  ACGGTGTTTT GGTCGAAGTC GATCAGTTGC AGGGTTTGGC TGCTCTCCTT
  51  GATGAACTCA CACAGCGGCT CGTAGAAGGC CTTTTGCTTG TGGTTCTCCC

primer31 Length: 100
   1  GACTTCGACC AAAACACCGT TAGCGACGAC AGCATCAAGA AGATCCTGGA
  51  GAGCATCAGC GAGAGCAAGT ACCACCACTA CCTGCGTCTG AACCCGAGCC

primer32   Length: 100
   1  TAACCAGAAC GCCCTTCTCG TCACAAGCCT TCAGCAGCTC TTGGATCTCC
  51  TCGTTCTCGC TCTTGATCAG GCTGCTGCTT TGGCTCGGGT TCAGACGCAG

primer33   Length: 100  .
   1  CGAGAAGGGC GTTCTGGTTA AGGCTTACTA CAAGTTCCCG CTGTGTCTGC
  51  CGACCGGCAC CTACTACGAC TACAACAGCG ACCGTTGGTG AGAGCTCCAC

primer34 Length: 62
   1  CCCCAAGCTT CCCGGGACTA GTTCTAGAGC GGCCGCCACC GCGGTGGAGC
  51  TCTCACCAAC GG
```

FIGURE 7

PRIMER SET 4

F1 Length: 82
1 GGGCGGATCC ATGGAGATCG AGAACAACCA AGCTCAACAA CCGAAGGCTG
51 AGAAGCTGTG GTGGGAGCTG GAGCTGGAGA TG

F2 Length: 100
1 CTGGTTAACG TTACCGCTGC TTGTCTGCTG CAAGAGGGCA GCTACTACCA
51 AGACAAGGAC GAGCGTCGTT ACATCATCAC CAAGGCTCTG CTGGAGGTTG

F3 Length: 100
1 CCGTACCACC ACCAACTACA TCGTTGCTTT CTGTGTTGTT CACAAGAACA
51 CCCAACCGTT CATCGAGAAG TACTTCAACA AGGCTGTTCT GCTGCCGAAC

F4 Length: 100
1 CAAGAACCTG TACCTGGACC GTATCCTGAG CCAAGATATC CGTAAGGAGC
51 TGACCTTCCG TAAGTGTCTG CAACGTTGTG TTCGTAGCAA GTTCAGCGAG

F5 Length: 100
1 CCGTTACCTG AGCGTTACCA ACAAGCAAAA GTGGGACCAA ACCAAGAAGA
51 AGCGTAAGGA GAACCTGCTG ACCAAGCTGC AAGCTATCAA GGAGAGCGAG

F6 Length: 100
1 GAAGCCGGCC GTTATGAAGA AGATCGCTAA GCGTCAAAAC GCTATGAAGA
51 AGCACATGAA GGCTCCGAAG ATCCCGAACA GCACCCTGGA GAGCAAGTAC

F7 Length: 100 .
1 CAAGATCCTG GGCAAGAAGT ACCCGAAGAC CGAGGAGGAG TACAAGGCTG
51 CTTTCGGCGA CAGCGCTAGC GCTCCGTTCA ACCCGGAGCT GGCTGGCAAG

F8 Length: 100 ..
1 CTGAGGTTTG GGACAACCTG ATCAGCAGCA ACCAACTGCC GTACATGGCC
51 ATGCTGCGTA ACCTGAGCAA CATCCTGAAG GCTGGCGTTA GCGACACCAC

F9 Length: 100
1 CCGCTGCAAT TCTTCAGCGC TATCGAGGCT GTTAACGAGG CGGTTACCAA
51 GGGCTTCAAG GCTAAGAAGC GTGAGAACAT GAACCTGAAG GGCCAAATCG

R1 Length: 97
1 GCAGCGGTAA CGTTAACCAG GTATTGCTTC GGGTCGTCGA TCTTAACACG
51 AACTTGGATG TCGTTTTGGT TCTCTTGCAT CTCCAGCTCC AGCTCCC

R2 Length: 100
1 GTAGTTGGTG GTGGTACGGA TGTACAGCTC GTTACGGATG TAAACAGCCA
51 GTTGACAGAT GAACTCCGGG TCGCTCTCAG CAACCTCCAG CAGAGCCTTG

R3 Length: 99
1 CGGTCCAGGT ACAGGTTCTT GAACTCGGTA GCGTCGAAGA TGTACAGAAC
51 TTGAGCGAAC TCACAAACCT CCAGCAGGTC GTTCGGCAGC AGAACAGCC

R5 Length: 99 .
1 CTTCATAACG GCCGGCTTCA GAGCCTTGAT AGCGTCCTCA ACGTTCATGA
51 TGTCGCCGGT CTCACGCTTG CTCTTGTCCT CGCTCTCCTT GATAGCTTG

FIGURE 8A

R6 Length: 100
1 GTACTTCTTG CCCAGGATCT TGTAAACACG TTCCTTCGGC TCGCTGATGT
51 GACAGAACTT GATCAGGTCC TTGAAGGTCA GGTACTTGCT CTCCAGGGTG

R7 Length: 99
1 CAGGTTGTCC CAAACCTCAG CGGTGTTGCC CTTAGCGCTC AGCTCGTTCT
51 CCCAGGTCTT GCTGATCTCG ATCTTCATAC GCTTGCCAGC CAGCTCCGG

R8 Length: 99
1 CGCTGAAGAA TTGCAGCGGG AACATCTTGC TGTTCTCAAC AGCCTTCGGC
51 TCACAGATCT TGTTGATAAC GATGCTGTGG GTGGTGTCGC TAACGCCAG

R9 Length: 101
1 CGAATTCGCC CTCCTCGGTT TGCTCCAGCT CCATGTCCTT CTTCTCCTCG
51 TCGGTCTTCT CAACAACCTC CTTAACAGCC TCGATTTGGC CCTTCAGGTT
101 C

FIGURE 8B

PRIMER SET 5

```
F10  Length: 96
  1  CCGAGGAGGG CGAATTCGTT AAGGTTAACG AGGGCATCGG CAAGCAATAC
 51  ATCAACAGCA TCGAGCTGGC TATCAAGATC GCTGTGAACA AGAACC

F11  Length: 99
  1  CATGAGCGGC GGCGCTAAGA AGTACGGCAG CGTTCGTACC TGTCTGGAGT
 51  GTGCTCTGGT TCTGGGCCTG ATGGTTAAGC AACGTTGTGA GAAGAGCAG

F12 Length: 100
   1  CGGGCGACGA GCTGCGTCCG AGCATGCAAA AGCTGCTGCA AGAGAAGGGC
  51  AAGCTGGGCG GCGGCACCGA CTTCCCGTAC GAGTGTATCG ATGAGTGGAC

F13   Length: 100
   1  CTACAGCGAC ATCAACGTTC GTGGCAGCAG CATCGTTAAC AGCATCAAGA
  51  AGTACAAGGA CGAGGTTAAC CCGAACATCA AAATCTTCGC TGTTGACCTG

F14  Length: 85
  1  CAAAATCTTC GGCATGAGCG ACAGCATCCT GAAGTTCATC AGCGCTAAGC
 51  AAGGCGGCGC TAACATGGTG GAGGTGATCA AGAAC

R10    Length: 100
   1  CTTAGCGCCG CCGCTCATGC TGGTGCTCAT GCTGCCGCTG ACGTCGCTGA
  51  AGATAGCGGT GTGGCCCTTG ATCTCGTCCA GGTTCTTGTT CACAGCGATC

R11   Length: 100
   1  GACGCAGCTC GTCGCCCGGC AGGTCAACCT CCAGGTAACA CTTGTTACAT
  51  TGGCTGCTCG GGCTGCTGAA GATGTAGAAG CTGCTCTTCT CACAACGTTG

R12   Length: 100
   1  GAACGTTGAT GTCGCTGTAG CCCTCAGCGA TCATCATGTC GCTCAGGATA
  51  ACGATGTTGT CAACGTGGGT CTTGTTCTTG GTCCACTCAT CGATACACTC

R13   Length: 98
  1  CGCTCATGCC GAAGATTTTG ATGTAGTTGT TCTCGTTGAA CTCGTCGCCC
 51  AGGTTCAGAC ACTTGCCGTA GCCCTCCAGG TCAACAGCGA AGATTTTG

R14    Length: 85  .
  1  GGGCGGTACC AAGCTTTCTA GACTAGTCTG CAGTCACTTT TGGCCGATCT
 51  TTTGCAGAGC GAAGTTCTTG ATCACCTCCA CCATG
```

FIGURE 9

GGGCGGATCCATGGAGATCGAGAACAACCAAGCTCAACAACCGAAGGCTGAGAAGCTGTGG
TGGGAGCTGGAGCTGGAGATGCAAGAGAACCAAAACGACATCCAAGTTCGTGTTAAGATCG
ACGACCCGAAGCAATACCTGGTTAACGTTACCGCTGCTTGTCTGCTGCAAGAGGGCAGCTA
CTACCAAGACAAGGACGAGCGTCGTTACATCATCACCAAGGCTCTGCTGGAGGTTGCTGAG
AGCGACCCGGAGTTCATCTGTCAACTGGCTGTTTACATCCGTAACGAGCTGTACATCCGTA
CCACCACCAACTACATCGTTGCTTTCTGTGTTGTTCACAAGAACACCCAACCGTTCATCGA
GAAGTACTTCAACAAGGCTGTTCTGCTGCCGAACGACCTGCTGGAGGTTTGTGAGTTCGCT
CAAGTTCTGTACATCTTCGACGCTACCGAGTTCAAGAACCTGTACCTGGACCGTATCCTGA
GCCAAGATATCCGTAAGGAGCTGACCTTCCGTAAGTGTCTGCAACGTTGTGTTCGTAGCAA
GTTCAGCGAGTTCAACGAGTACCAACTGGGCAAGTACTGTACCGAGAGCCAACGTAAGAAG
ACCATGTTCCGTTACCTGAGCGTTACCAACAAGCAAAAGTGGGACCAAACCAAGAAGAAGC
GTAAGGAGAACCTGCTGACCAAGCTGCAAGCTATCAAGGAGAGCGAGGACAAGAGCAAGCG
TGAGACCGGCGACATCATGAACGTTGAGGACGCTATCAAGGCTCTGAAGCCGGCCGTTATG
AAGAAGATCGCTAAGCGTCAAAACGCTATGAAGAAGCACATGAAGGCTCCGAAGATCCCGA
ACAGCACCCTGGAGAGCAAGTACCTGACCTTCAAGGACCTGATCAAGTTCTGTCACATCAG
CGAGCCGAAGGAACGTGTTTACAAGATCCTGGGCAAGAAGTACCCGAAGACCGAGGAGGAG
TACAAGGCTGCTTTCGGCGACAGCGCTAGCGCTCCGTTCAACCCGGAGCTGGCTGGCAAGC
GTATGAAGATCGAGATCAGCAAGACCTGGGAGAACGAGCTGAGCGCTAAGGGCAACACCGC
TGAGGTTTGGGACAACCTGATCAGCAGCAACCAACTGCCGTACATGGCCATGCTGCGTAAC
CTGAGCAACATCCTGAAGGCTGGCGTTAGCGACACCACCCACAGCATCGTTATCAACAAGA
TCTGTGAGCCGAAGGCTGTTGAGAACAGCAAGATGTTCCCGCTGCAATTCTTCAGCGCTAT
CGAGGCTGTTAACGAGGCGGTTACCAAGGGCTTCAAGGCTAAGAAGCGTGAGAACATGAAC
CTGAAGGGCCAAATCGAGGCTGTTAAGGAGGTTGTTGAGAAGACCGACGAGGAGAAGAAGG
ACATGGAGCTGGAGCAAACCGAGGAGGGCGAATTCGTTAAGGTTAACGAGGGCATCGGCAA
GCAATACATCAACAGCATCGAGCTGGCTATCAAGATCGCTGTGAACAAGAACCTGGACGAG
ATCAAGGGCCACACCGCTATCTTCAGCGACGTCAGCGGCAGCATGAGCACCAGCATGAGCG
GCGGCGCTAAGAAGTACGGCAGCGTTCGTACCTGTCTGGAGTGTGCTCTGGTTCTGGGCCT
GATGGTTAAGCAACGTTGTGAGAAGAGCAGCTTCTACATCTTCAGCAGCCCGAGCAGCCAA
TGTAACAAGTGTTACCTGGAGGTTGACCTGCCGGGCGACGAGCTGCGTCCGAGCATGCAAA
AGCTGCTGCAAGAGAAGGGCAAGCTGGGCGGCGGCACCGACTTCCCGTACGAGTGTATCGA
TGAGTGGACCAAGAACAAGACCCACGTTGACAACATCGTTATCCTGAGCGACATGATGATC
GCTGAGGGCTACAGCGACATCAACGTTCGTGGCAGCAGCATCGTTAACAGCATCAAGAAGT
ACAAGGACGAGGTTAACCCGAACATCAAAATCTTCGCTGTTGACCTGGAGGGCTACGGCAA
GTGTCTGAACCTGGGCGACGAGTTCAACGAGAACAACTACATCAAAATCTTCGGCATGAGC
GACAGCATCCTGAAGTTCATCAGCGCTAAGCAAGGCGGCGCTAACATGGTGGAGGTGATCA
AGAACTTCGCTCTGCAAAAGATCGGCCAAAAGTGACTGCAGACTAGTCTAGAAAGCTTGGT
ACCGCCC

FIGURE 10

GGGGCCATGGATGAGCCGTCGTAACCAAAAGAAGCCGCAAGCTCCGATCGGCAACGAGACC
AACCTGGACTTCGTTCTGCAAAACCTGGAGGTTTACAAGAGCCAAATCGAGCACTACAAGA
CCCAACAACAACAAATCAAGGAGGAGGACCTGAAGCTGCTGAAGTTCAAGAACCAAGACCA
AGACGGCAACAGCGGCAACGACGACGACGACGAGGAGAACAACAGCAACAAGCAACAAGAG
CTGCTGCGTCGTGTTAACCAAATCAAGCAACAAGTTCAACTGATCAAGAAGGTTGGCAGCA
AGGTTGAGAAGGACCTGAACCTGAACGAGGACGAGAACAAGAAGAACGGCCTGAGCGAGCA
ACAAGTTAAGGAGGAGCAACTGCGTACCATCACCGAGGAGCAAGTTAAGTACCAAAACCTG
GTTTTCAACATGGACTACCAACTGGACCTGAACGAGAGCGGCGGCCACCGTCGTCACCGTC
GCGAGACCGACTACGACACCGAGAAGTGGTTCGAGATCAGCCACGACCAAAAGAACTACGT
TAGCATCTACGCTAACCAAAAGACCAGCTACTGTTGGTGGCTGAAGGACTACTTCAACAAG
AACAACTACGACCACCTGAACGTTAGCATCAACCGTCTGGAGACCGAGGCTGAGTTCTACG
CTTTCGACGACTTCAGCCAAACCATCAAGCTGACCAACAACAGCTACCAAACCGTTAACAT
CGACGTCAACTTCGACAACAACCTGTGTATCCTGGCTCTGCTGCGTTTCCTGCTGAGCCTG
GAGCGTTTCAACATCCTGAACATCCGTAGCAGCTACACCCGTAACCAATACAACTTCGAAA
AGATCGGCGAGCTGCTGGAGACCATCTTCGCTGTTGTTTTCAGCCACCGTCACCTGCAAGG
CATCCACCTGCAAGTTCCGTGTGAGGCTTTCCAATACCTGGTTAACAGCAGCAGCCAAATC
AGCGTTAAGGACAGCCAACTGCAAGTTTACAGCTTCAGCACCGACCTGAAGCTGGTTGACA
CCAACAAGGTTCAAGACTACTTCAAGTTCCTGCAAGAGTTCCCGCGTCTGACCCACGTGAG
CCAACAAGCTATCCCGGTTAGCGCTACCAACGCTGTTGAGAACCTGAACGTTCTGCTGAAG
AAGGTTAAGCACGCTAACCTGAACCTGGTTAGCATCCCGACCCAATTCAACTTCGACTTCT
ACTTCGTTAACCTGCAACACCTGAAGCTGGAGTTCGGCCTGGAGCCGAACATCCTGACCAA
GCAAAAGCTGGAGAACCTGCTGCTGAGCATCAAGCAAAGCAAGAACCTGAAGTTCCTGCGT
CTGAACTTCTACACCTACGTTGCTCAAGAGACCAGCCGTAAGCAAATCCTGAAGCAAGCTA
CCACCATCAAGAACCTGAAGAACAACAAGAACCAAGAGGAGACTCCGGAGACCAAGGACGA
GACCCCGAGCGAGAGCACCAGCGGCATGAAGTTCTTCGACCACCTGAGCGAGCTGACCGAG
CTGGAGGACTTCAGCGTTAACCTGCAAGCTACCCAAGAGATCTACGACAGCCTGCACAAGC
TGCTGATCCGTAGCACCAACCTGAAGAAGTTCAAGCTGAGCTACAAGTACGAGATGGAGAA
GAGCAAGATGGACACCTTCATCGATCTGAAGAACATCTACGAGACCCTGAACAACCTGAAG
CGTTGTAGCGTTAACATCAGCAACCCGCACGGCAACATCAGCTACGAGCTGACCAACAAGG
ACAGCACCTTCTACAAGTTCAAGCTGACCCTGAACCAAGAGCTGCAACACGCTAAGTACAC
CTTCAAGCAAAACGAATTCCAATTCAACAACGTTAAGAGCGCTAAGATCGAGAGCAGCAGC
CTGGAGAGCCTGGAGGACATCGACAGCCTGTGTAAGAGCATCGCCAGCTGTAAGAACCTGC
AAAACGTTAACATCATCGCTAGCCTGCTGTACCCGAACAACATCCAAAAGAACCCGTTCAA
CAAGCCGAACCTGCTGTTCTTCAAGCAATTCGAGCAACTGAAGAACCTGGAGAACGTTAGC
ATCAACTGTATCCTGGACCAACACATCCTGAACAGCATCAGCGAGTTCCTGGAGAAGAACA
AGAAGATCAAGGCTTTCATCCTGAAGCGTTACTACCTGCTGCAATACTACCTGGACTACAC
CAAGCTGTTCAAGACCCTGCAACAACTGCCGGAGCTGAACCAAGTTTACATCAACCAACAA
CTGGAGGAGCTGACCGTTAGCGAGGTTCACAAGCAAGTTTGGGAGAACCACAAGCAAAAGG
CCTTCTACGAGCCGCTGTGTGAGTTCATCAAGGAGAGCAGCCAAACCCTGCAACTGATCGA
CTTCGACCAAAACACCGTTAGCGACGACAGCATCAAGAAGATCCTGGAGAGCATCAGCGAG
AGCAAGTACCACCACTACCTGCGTCTGAACCCGAGCCAAAGCAGCAGCCTGATCAAGAGCG
AGAACGAGGAGATCCAAGAGCTGCTGAAGGCTTGTGACGAGAAGGGCGTTCTGGTTAAGGC
TTACTACAAGTTCCCGCTGTGTCTGCCGACCGGCACCTACTACGACTACAACAGCGACCGT
TGGTGAGAGCTCCACCGCGGTGGCGGCCGCTCTAGAACTAGTCCCGGGAAGCTTGGGG

FIGURE 11